The Dissertation Committee for James Peter Monaco
certifies that this is the approved version of the following dissertation:

Active Binocular Vision: Phase-Based Registration and Optimal Foveation

Committee:

Alan C. Bovik, Supervisor

Lawrence Cormack, Supervisor

J.K. Aggarwal

Joydeep Ghosh

Ross Baldick

Active Binocular Vision: Phase-Based Registration and Optimal Foveation

by

James Peter Monaco, B.S.; M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN
May 2007

Dedicated to my family.

Acknowledgments

I would foremost like to thank my family—James, Linda, and Jason Monaco—for their support in this long journey. I am sure that without their help I would have had no chance of succeeding. I would also like to thank my friends for their support, especially Vicky Zipper. This process requires sacrifices not only from the students, but also from those who support them.

I would like to thank Dr. Alan Bovik for his help and encouragement. Dr. Bovik not only helped me with ideas and support, but also was instrumental in my return to Austin after a substantial hiatus. Without his sound advice I would never have finished. Nor would I have learned to appreciate Chai tea.

Thanks also to the past and current members of the Laboratory for Image and Video Engineering, especially Umesh Rajashekar. Their advice and knowledge made this process far easier than it could have been.

Thanks to Brent Wells and Bob Gibbons at Raytheon. Their guidance and friendship were irreplaceable. Much appreciation also goes to Jeff Wehnes and VuCOMP. Without the support of these companies and persons, this endeavor would not have been possible.

Active Binocular Vision: Phase-Based Registration and Optimal Foveation

Publication No. ____________________

James Peter Monaco, Ph.D.
The University of Texas at Austin, 2007

Supervisors: Alan C. Bovik
Lawrence Cormack

Active binocular vision systems are powerful tools in machine vision. With a virtually unlimited field-of-view they have access to huge amounts of information, yet are able to confine their resources to specific regions of interest. Since they can dynamically interact with the environment, they are able to successfully address problems that are ill-posed to passive systems. A primary goal of an active binocular vision systems is to ascertain depth information. Since they employ two cameras and are able two sample a scene from two distinct vantage points, they are well suited for such a task.

The depth recovery process is composed of two interrelated components: image registration and sampling. Image registration is the process of determining corresponding points between the stereo images. Once points in the images have been matched, 3D information can be recovered via triangulation. Image sampling determines how the image is discretized and represented. Image

registration and sampling are highly interdependent. The choice of sampling scheme can profoundly impact the accuracy and complexity of the registration process. In many situations, particular registration algorithms are simply incompatible with some sampling schemes.

In this dissertation we meticulously address both registration and sampling in the context of stereopis for active binocular vision systems. Throughout the development of this work, contributions in each area are addressed with the an eye toward their eventual integration into a cohesive registration procedure appropriate for active binocular vision systems. The actual synthesis is a daunting task that is beyond the scope of this single dissertation. The focus of this work is to assiduously analyze both registration and sampling, establishing a solid foundation for their future aggregation.

One of the most successful approaches to image registration is phase-differencing. Phase-differencing algorithms provide a fast, powerful means for depth recovery. Unfortunately, phase-differencing techniques suffer from two significant impediments: phase nonlinearities and neglect of multispectral information. This dissertation uses the amenable properties of white noise images to analytically quantify the behavior of phase in these regions of phase nonlinearity. The improved understanding gained from this analysis enables us to create a new, more effective method for identifying these regions based on the second derivative of phase. We also suggest a novel approach that combines our method of nonlinear phase detection with strategies of both phase-differencing and local correlation. This hybrid approach retains the advantageous properties of phase-differencing while incorporating the multi-spectral aspects of local correlation.

This task of registration is greatly simplified if the camera geometry is known and the search for corresponding points can be restricted to epipolar lines. Unfortunately, computation of epipolar lines for an active system requires calibration which can be both highly complex and inaccurate. While it is possible to register images without calibration information, such unconstrained algorithms are usually time consuming and prone to error. In this dissertation we propose a compromise. Even without the instantaneous knowledge of the system geometry, we can restrict the region of correspondence by imposing limits on the possible range of configurations, and as a result, confine our search for matching points to what we refer to as *epipolar spaces.* For each point in one image, we define the corresponding epipolar space in the other image as the union of all associated epipolar lines over all possible system geometries. Epipolar spaces eliminate the need for calibration at the cost of an increased search region. Since the average size of a search space is directly related to the accuracy and efficiency of any registration algorithm, it is essential to mitigate this increase. The major contribution of this dissertation is the derivation of an optimal nonuniform sampling strategy that minimizes the average area per epipolar space.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The introduction is divided into three sections. The first section provides a brief background on the topics to be discussed. The next section delineates the contributions of this dissertation. The final section provides an overview of the remaining chapters of the dissertation.

1.1 Background

This section begins with a brief background of the topics addressed in this dissertation. First, it discusses active binocular vision systems. Second, the process of image registration is examined. Next, it covers epipolar geometry. Finally, an overview of image sampling is provided.

1.1.1 Active Binocular Vision Systems

An active binocular vision system (ABVS) is a powerful tool in machine vision. As the name implies, the system is binocular: it has two cameras; the system is active: it has the ability to dynamically alter its configuration in response to the specific environment. For example, most ABVS can rotate and displace each camera independently. Additionally, intrinsic camera parameters such as focal length can be constantly modified. Some ABVS are even integrated as components of mobile robots. Fig. 1.1 provides an example of an ABVS designed by Klarquist and Bovik [70] called FOVEA.

Figure 1.1: FOVEA - Active binocular vision system by Klarquist and Bovik.

Active binocular vision systems provide a effective means for extracting information from a complex scene. With a virtually unlimited field-of-view they have access to huge amounts of information, yet are able to confine their resources to specific regions of interest. Since they are able to dynamically interact with the environment, they can successfully address problems in computational vision that are ill-posed to a passive system [4]. The term “ill-posed” describes problems for which one of more of the following is true [55]:

1. A solution does not exist.

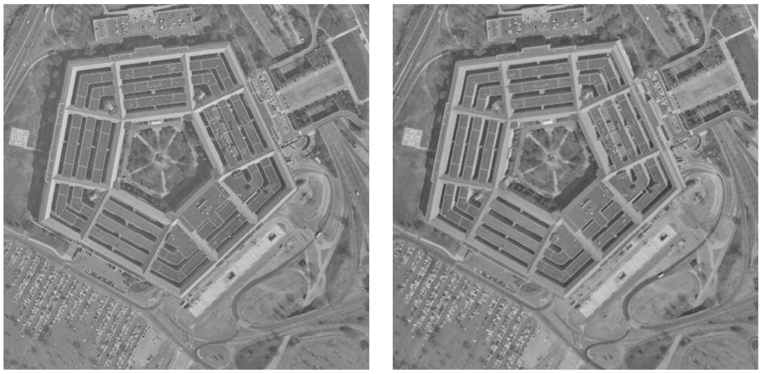


Figure 1.2: Stereo images of pentagon.

2. The solution is not unique.
3. The solution does not depend continuously on the initial data.

A primary goal of an ABVS is to ascertain depth information. Since it employs two cameras and is able two sample a scene from two distinct vantage points, an AVBS is especially well suited for such a purpose. Fig. 1.2 shows a pair of stereo images, i.e. images of the same scene from two distinct points of view. Using these images and the geometry of camera system, the depth of the objects in the scene can be recovered via triangulation. To understand this, consider Fig. 1.3 which diagrams a typical stereo configuration. The cameras are pinhole projection models. A pinhole projection model assumes that every incident light ray must pass through the pinhole before contacting the image plane. The points C_l and C_r indicate the locations of the pinholes of the left and right image planes, respectively. The 3D point P projects onto the left and right image planes at the points I_l and I_r, respectively. Therefore, if we are able to determine that I_l and I_r are projections of the same 3D point P we can recover the location of P by simple triangulation.

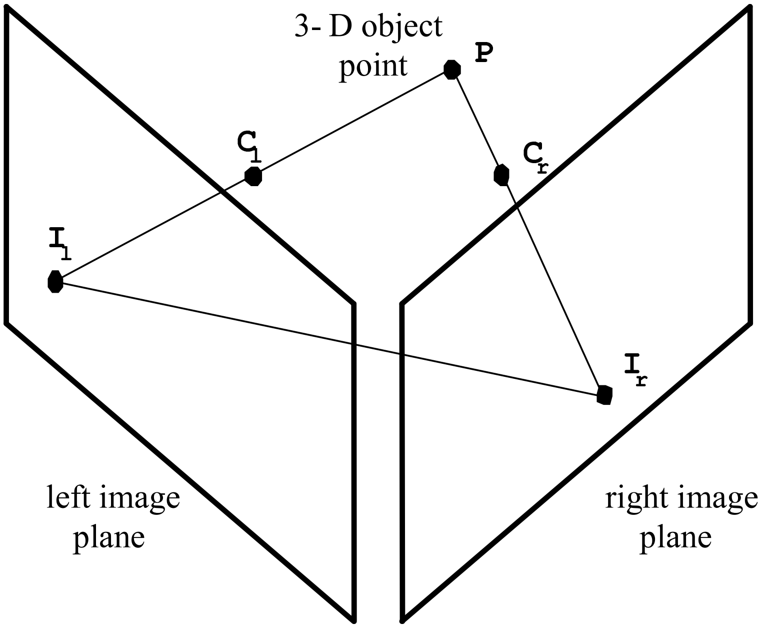


Figure 1.3: Depth recovery via triangulation.

The process of matching points between images is known as the "correspondence problem" [76]. Techniques that solve the correspondence problem are known as registration algorithms. Using an registration algorithm developed in this dissertation the images in Fig. 1.2 were matched, enabling the production of the depth map in Fig. 1.4.

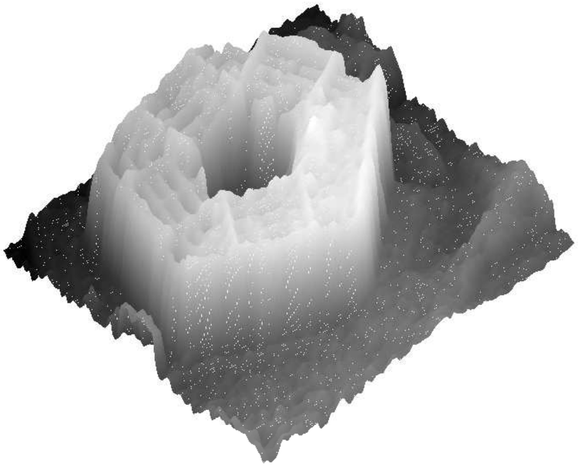


Figure 1.4: Depth map of pentagon.

1.1.2 Image Registration

At first glance the correspondence problem may seem rather trivial. It is, in fact, one of the most difficult and studied problems in computer vision. The difficulties arise from several factors. First, inherent variations between cameras result in differences in lighting, contrast, and noise. Second, disparate vantage points cause foreshortening, the shrinking or expansion of objects due to prospective projection, and often result in occlusions. For such situations there can be no exact correspondence between the two views; and the problem is ill-posed. Fortunately, the problem can become well-posed, yielding a unique solution, if the binocular camera system is active [4].

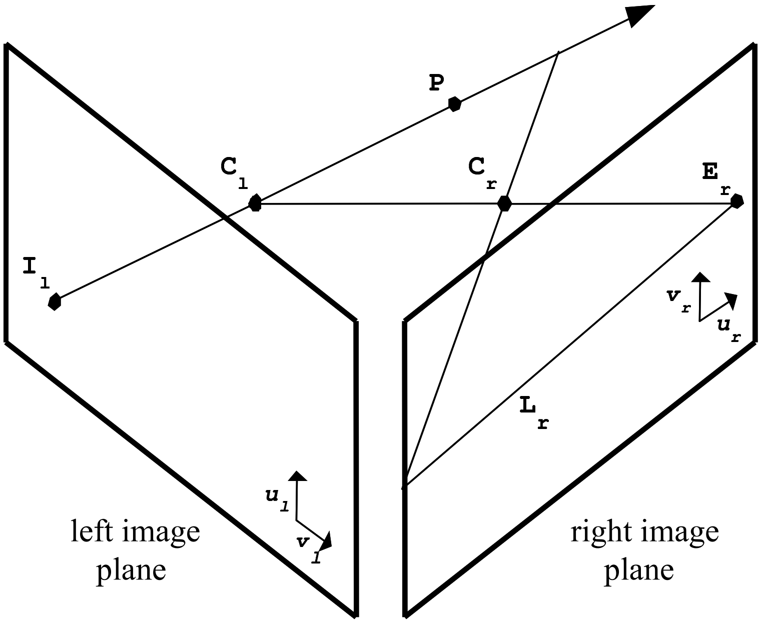


Figure 1.5: Epipolar geometry.

1.1.2.1 Epipolar Geometry

One way to significantly simplify the correspondence problem is take advantage the geometry of the cameras. If the exact configuration of the system is known then matching points are restricted to lines called *epipolar lines*. That is, given a point in the left image, if we want to identify the matching point in the right then we need only search along the appropriate line in the right image (or vice versa). Fig. 1.5 illustrates the principle of epipolar geometry. We know that any 3D point P that projects onto the point I_l in the left image plane must lie somewhere on the 3D line $I_l\vec{C}_l$. Projecting this 3D line onto the right image plane produces the line L_r. The line L_r is the epipolar line associated with the point I_l.

1.1.2.2 Registration Algorithms

There are two primary approaches to stereo image registration. The first approach advocates matching physically meaningful features such as edges or corners [39]. Such an approach has certain advantages: the features are relatively unambiguous, insensitive to changes in lighting and contrast, and tend to occur at locations in the image that are important to us. Unfortunately, they appear only sparsely, allowing depth recovery at only a few locations in the image. Using such techniques it is difficult to recover depth information over entire scenes.

The second methodology suggests using image properties that can be measured at every point in the image. These measurements tend to have more of a mathematical framework and less of an intuitive physical interpretation. For example, consider the following rudimentary algorithm: given a pixel in the left image, search for the pixel in the right that has the most similar intensity. (A more robust algorithm, instead of performing a point by point comparison, would examine image patches.) Since similarity measures can be made between all point in both images, a dense depth map can be constructed.

Another image property available at every point in an image is local phase. Local phase has been employed as the primary component in a large array of registration algorithms [25, 62, 99]. The concept of local phase is best understood through illustration. Consider Fig. 1.6. To measure the local phase at a point I, an appropriately sized image patch centered about I is compared with the two wavelike patterns in Fig. 1.6. These comparison operations produce values indicating how similar the image patch is to each pattern. The ratio of these similarity values can be considered the local phase at that point. A robust registration algorithm using local phase would employ multiple pairs of patterns like the ones in Fig. 1.6 but tuned to different sizes, frequencies, and orientations. In fact, studies suggest that some biological visual systems (including the human visual system) may use similar mechanisms [38, 65].

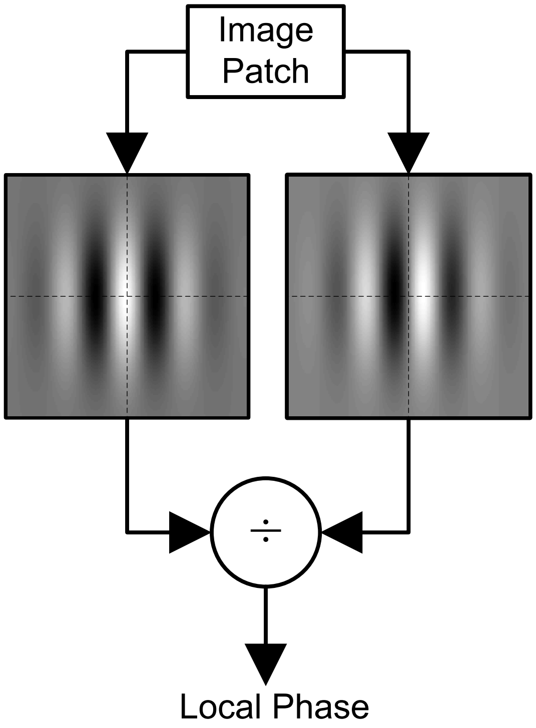


Figure 1.6: Computation of local phase.

1.1.3 Image Sampling

The image projected onto each camera's imaging plane is continuous. For processing with a computer, this continuous signal signal must be discretized. In almost all cases, discretization is performed by uniformly sampling the continuous image in a rectangular grid-like pattern. The pentagon images shown in Fig. 1.2 are examples of uniformly sampled images. While such a sampling strategy is appropriate for our viewing, it may not be the best representation for computational vision tasks. In fact, the human visual system uses a very different approach when sampling and storing visual information.

The human eye samples a scene more densely toward the central region (fovea) than in the periphery [53]. Fig. 1.7 diagrams the density of photorecep-

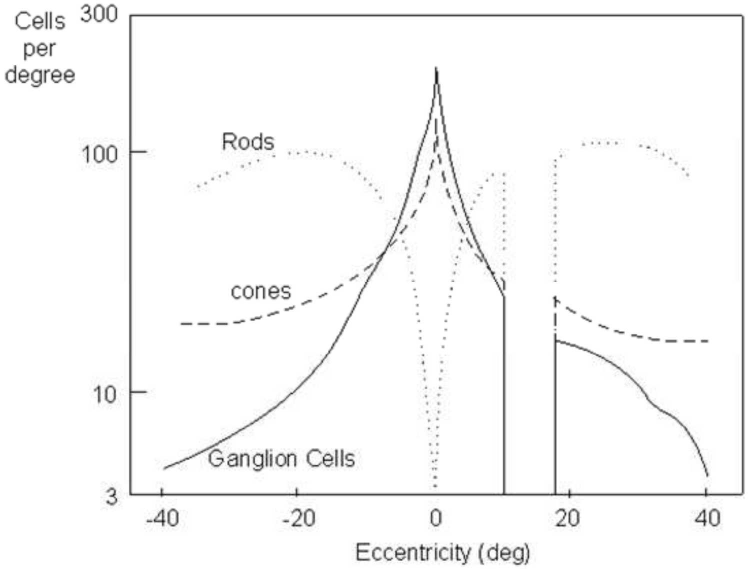


Figure 1.7: Density of photoreceptors for the human visual system.

tors in the human eye. This figure illustrates that the majority of receptors are allocated toward the central region of the eye. The density rapidly diminishes with increasing eccentricity.

A model of this nonuniform retinal sampling is shown in Fig. 1.8. Each enclosed area represents one unique sample. Notice how the resolution decreases with the distance from the periphery. The cortical mapping, how each sample is encoded in the brain, is also illustrated in Fig. 1.8. The circles in the retinal coordinates are mapped to straight lines in the cortical space. This transformation from retinal to cortical coordinates is called a log-polar mapping and was first proposed by Schwartz [105–107] as a model for the human visual system. This mapping mathematically simplifies operations such as dilation and rotation, which may explain why humans perform so well at many pattern recognition tasks [128].

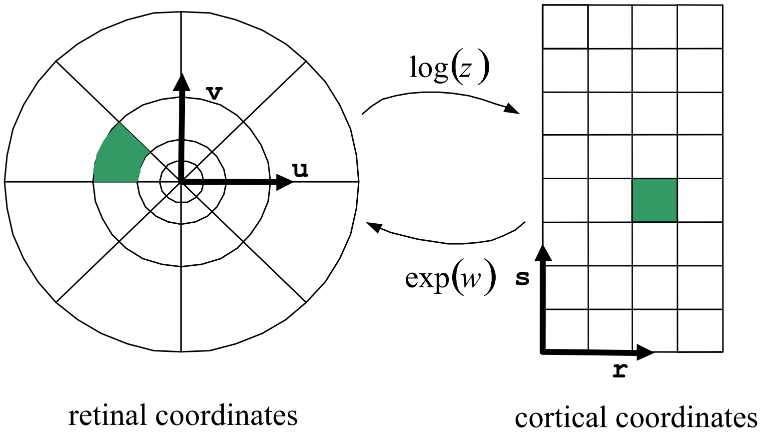
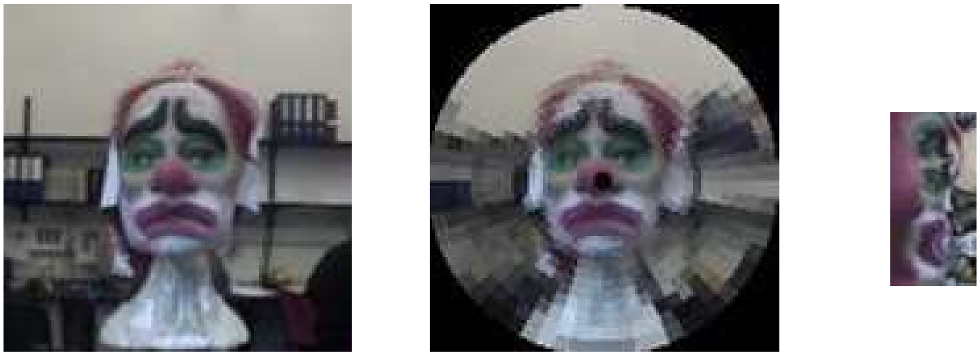


Figure 1.8: Retinal and cortical mappings of the human visual systems proposed by Schwartz.

Fig. 1.9 illustrates three different image representations. The first image shows a uniformly sampled image of a clown. The second image demonstrates the retinal mapping of the clown image. The third image illustrates the cortical mapping. Obviously, the cortical image is difficult to interpret visually. However, for tasks as such as registration and pattern recognition, it may be superior.



uniform sampling retinal cortical

Figure 1.9: Different image representations (LiraLab).

1.2 Contributions

Image sampling and registration are two integral and interrelated components in the depth recovery procedure of an ABVS. This dissertation provides an advancement in both areas. These advancements were developed with an eye toward their eventual integration into a cohesive registration procedure appropriate for ABVS.

The next sections briefly outline the contributions of this dissertation. The first section discusses advancements in phase-based registration. The next section introduces the concept of epipolar spaces, which are the extension of epipolar lines to the realm of active binocular vision. The final section discusses optimal sampling strategies designed specifically to optimize certain properties of these epipolar spaces.

1.2.1 Advances in the Understanding and Application of Local Phase for Registration

There are certain limitations associated with the use of local phase for image registration. Under some fairly common circumstances phase measurements becomes very poor, making them unusable for matching. Furthermore, combining local phase measures from many different wave patterns is not a simple task.

Authors have worked toward determining the reasons that phase fails in certain regions [19,48]. Unfortunately, the methods for detecting these regions are rather ad hoc. It is the purpose of this dissertation is to introduce a new paradigm that better elucidates these regions. This paradigm provides a firm theoretical foundation that explains the success of the *ad hoc* methods, while also improving upon them.

Combining local phase measurements to produce a single best correspondence between images is a complex procedure that has never been conclusively addressed. Most approaches combine these phase measures using heuristic procedures that negate many of the advantageous properties of local phase. Again this dissertation introduces a new paradigm that provides a means for combining phase values in an intuitive, theoretically sound, and algorithmically effective manner.

1.2.2 Epipolar Spaces

Using epipolar constraints requires an exact knowledge of the system geometry. This is obtained through a process called calibration. Unfortunately, calibrating an ABVS is highly complex and often inaccurate [111], making it highly undesirable if not untenable. Though, without calibration, we are unable to know the precise geometric configuration, we can *a priori* restrict the range of geometric configurations. If we restrict such parameters as camera rotation angles, focal length, camera separation distance, etc. we can still confine the possible locations of matching points. Since each unique geometric configuration produces a unique epipolar line, the search for matching points can be restricted to the union of all possible epipolar lines produced from all possible configurations. Fig. 1.10 illustrates the point I_r and the attendant epipolar lines resulting from nine different geometric configurations. In practice, the entire epipolar line is not considered when searching for a matching point. Usually the search is confined to some horizontal distance around the point I_r. This restriction is shown in Fig. 1.10 as parallel dashed lines.

The epipolar lines in Fig. 1.10 were constructed from nine disjoint, randomly selected geometries. If instead, the lines were drawn for a contin-

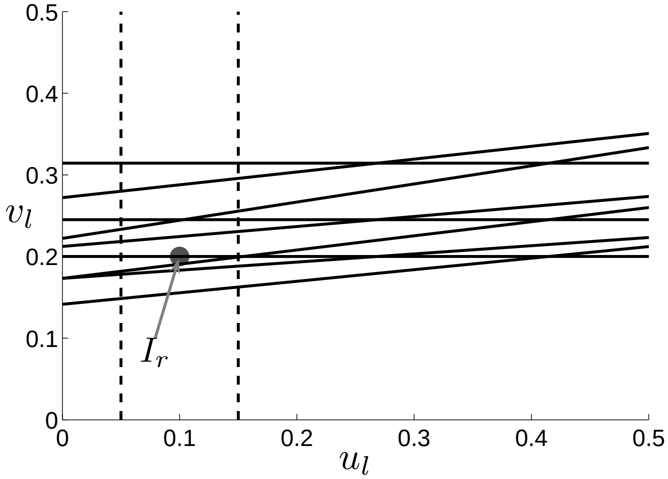


Figure 1.10: Left image epipolar lines corresponding to right image point I_r for several unique geometric configurations. The parallel dashed lines delimit a reasonable range around I_r in which to search for a matching point.

uous range of geometries they would merge into a continuous region. We call this continuous region an *epipolar space*. Several epipolar spaces are illustrated in Fig. 1.11. The spaces are confined horizontally to a reasonable range around their concomitant points. These horizontal bounds are represented with dashed lines. In summary, for an uncalibrated ABVS confined to a certain range of geometries a matching point is restricted to an epipolar space.

1.2.3 Optimal Sampling Strategies

Recognizing the advantages of nonuniform sampling methods such as log-polar, many authors have applied a wide range of sampling schemes to

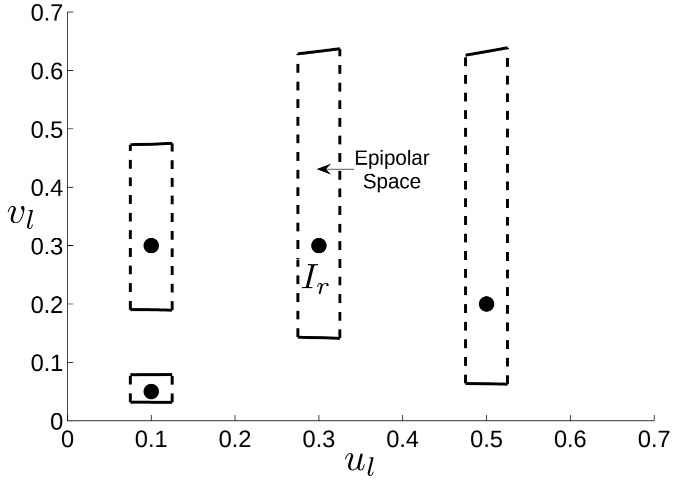


Figure 1.11: Various epipolar spaces. The large dots represent points from the right image. Each point is enclosed in its attendant epipolar space in the left image.

binocular visions tasks, resulting in various degrees of success. In most cases the foveation (i.e. sampling) strategies have been developed or ascertained elsewhere; they were not specifically designed for use in a mechanical ABVS. While sampling schemes, such as the log-polar retinal/cortical structure, may be extremely advantageous for biological systems employing biological sensors for biological tasks using biological registration algorithms it is likely that for our specific imaging devices, machine tasks, and registration algorithms such a mapping is not optimal. The foremost contribution of this dissertation is the derivation of an optimal strategy for a mechanical ABVS.

Before any optimal scheme can be constructed, we must first establish the criteria we wish to optimize. This is done by first recognizing that the computational complexity and accuracy of a registration algorithm is directly related to the number of possible points that have to be examined when searching for a match [108]. We define an optimal sampling scheme as a foveation strategy that minimizes the average number of points that must be considered when searching for matching points. For an uncalibrated ABVS this average is specified by the average size of an epipolar space. In this dissertation we derive the sampling scheme that minimizes the average area of an epipolar space.

1.3 Overview

The remainder of the dissertation is divided into four chapters. The following chapter provides a review of the current literature. The next chapter addresses phase-differencing for stereo registration. This is followed by a chapter concerning the derivation of epipolar spaces and their attendant optimal sampling schemes. The final chapter offers a conclusion and then discusses future research.

Chapter 2

Review of the Literature

2.1 Image Registration

In order to recover depth via triangulation a pair of stereo images must first be registered. Such correspondence algorithms can be divided into two categories: feature based and area based. Feature based algorithms [15, 39, 54, 67, 68, 76, 76] attempt to match the images at specific locations. Such locations often correspond to peaks or zero-crossings in the filtered images. These monocular tokens are assumed to represent such intuitive phenomena as edges or corners. Since such features occur sporadically in the images, any matching algorithm exclusively using them can only expect a sparse recovery of depth with respect to the entire field-of-view.

Area based algorithms compute and compare image properties that are available at every point in both images. The term "area based" refers to the fact that these properties are measured over some area surrounding the point in question. One of the earliest and most intuitive area based techniques is local cross-correlation. Unfortunately, cross-correlation has been shown to be sensitive to lighting, contrast, prospective projections, and noise [56], all of which can be common problems when images are taken from two distinct vantage points with two different cameras. Some of these difficulties can be mitigated by bandpass filtering the images prior to performing any correlation. Along this same vein come the most popular and powerful of the area based algorithms: phase-based correspondence techniques.

Phase-based algorithms, with techniques such as phase differencing and phase-correlation, have proven to be a powerful means for stereo image correspondence [15,23–26,61,64,99,115,131,133]. These techniques can produce disparity maps to sub-pixel accuracy, without requiring explicit sub-pixel reconstruction. Since relevant phase information is available at any point in the image, not only at specific tokens as in feature based methods, phase-based methods yield dense disparity maps. Additionally, phase is relatively insensitive to typical inter-image differences such as lighting, shadows, noise [35], and perspective deformations [49].

The premise of phase differencing methods originates in the Fourier Shift Theorem, which states that the Fourier Transforms of two signals related by a global shift are themselves identical up to a phase difference equal to the frequency times the shift. This linear relationship makes it possible to immediately recover this shift or disparity at any frequency (that is present in the signal) by simply dividing the phase difference by the frequency. Unfortunately, stereo images are not related by a global shift, but instead, are better modeled by many local shifts. Consequently, the generation of accurate disparity maps requires a spatially localized estimate of the phase. This can be accomplished by various means such as short-time Fourier transforms or wavelets. Because of its joint optimality in space and frequency, the Gabor function [52] is the most popular choice. In general, any localized quadrature filter pair can produce local phase measurements.

Evidence tends to suggest that biological systems may use bandpass filters to extract local phase information for stereo registration. The receptive fields of simple cells in the visual cortex can be well modeled by Gabor functions tuned to different frequencies and of different spatial extent [37, 66]. It has also been demonstrated that phase information is combined over different spatial scales [77] for stereopsis. A phase-shift model has been proposed in several papers [38, 82].

In [99], Sanger first used a phase differencing method for to stereo registration. Sanger extracted local phase at each image point using Gabor functions. He recovered the disparity by dividing the phase difference between the two views by the center frequency of the Gabor function. The disparity outputs for five Gabor functions spaced at 1/2 octave bandwidths were combined to produce a dense disparity map. Since such filters are only able to reliably detect disparities of 1/2 their wavelength due to phase-wrapping, he was only able produce a coarse disparity map. His algorithms also exhibited problems near disparity gradients.

In [48, 49, 61, 62, 64], Fleet, Jenkin, and Jepson expanded this approach in several ways. First, they noted that the constant of linearity relating phase and displacement could be better approximated by directly determining the first derivative of phase. This proved to be a better estimate than the center frequency of the Gabor function as proposed by Sanger. Second, they introduced a coarse-to-fine procedure, allowing larger scales to bring the images into approximate alignment before proceeding to finer scales. Third, they demonstrated that phase is robust with respect to image deformations caused by prospective projections. Finally, they studied the impact of filter localization upon this linear relationship. Perhaps the most comprehensive paper in this direction is [48]. Here Fleet *et. al.* models images as one-dimensional Gaussian white noise processes and derives bounds for the expected mean phase difference and absolute mean phase difference under the operations of translation and dilation.

From these seminal papers many others have expanded the work in important ways. A prevalent approach was to replace the Gabor function with another filter capable of gathering phase information. Methods such as the FFT [3, 129], complex wavelets [73, 133], image derivatives [88], and spatial Gaussians [41] were used to extract phase information in place of the Gabor function. Since there is still no widely accepted metric for comparing the quality of different stereo registration algorithms, it is difficult to assess whether one filter or another is superior. Another important advancement was provided by further examining the effects of localization. In [19], Cai examined the instability of localized phase by considering registration by phase differencing as a type of Newton iteration. He pointed out that the instability of phase measurements in certain regions violates the criteria necessary for convergence.

Others have addressed the shortcomings of phase differencing algorithms in the context of their dependence on coarse-to-fine processing schemes. Coarse-to-fine processing allows larger scales to bring the images into near registration before proceeding to smaller scales. At each scale disparity information is procured from a small range of frequencies. Images with relatively little energy in these frequency bands will tend to produce erroneous estimates that will further contaminate higher scales as the iteration proceeds. To mitigate this problem, alternative means for using phase information have been suggested. The goal of these techniques was to make phase measurements over the entire spectrum and then combine the results into one robust value. The difficulty in combining these values arises from the periodic nature of phase values. This periodicity can result in ambiguities in the disparity estimates.

One of the most robust solutions was proposed by Chen and Bovik in [24–26]. In these works the authors applied an AM-FM surface albedo model combining both multichannel Gabor demodulation techniques and coarse-to-fine processing. This unique combination proved an effective means for integrating information across multiple frequencies and scales while eliminating the usually attendant complication of phase-wrapping.

Other typical approaches applied short-time Fourier transformations [47, 94, 129] or wavelet decompositions [73, 133] and then globally searching between images for locations that minimize the difference in their coefficients. This difference is predominately a function of the local phase. These methods are time consuming, disregarding the linearity of phase that allows phase differencing to rapidly convergence to high resolution solutions. Furthermore, their resolution is inherently limited by the sampling of the images, and consequently, such algorithms require ancillary methods for any sub-pixel estimation. Other methodologies used phase differencing to realize disparity estimates at multiple frequencies then heuristically combined them employing secondary features such as magnitude information [87, 132]. Phase measurements have also been combined over multiple frequencies using probabilistic schemes [115], spatial representations [41], linear regression [2, 3] and image derivatives [88].

2.2 Sampling

The application of sampling strategies to stereo vision tasks is becoming more prevalent. Space-variant transformations such as log-polar [127], reciprocal wedge transform [71], foveate wavelet transform [125], and fisheye [110] have been successfully applied to binocular vision problems such as vergence [72, 74], time-to-impact analysis [114], and depth recovery of a scene [23, 72, 103]. The success of these sampling strategies has prompted the development of spatially-variant sensor arrays [7, 44, 45].

By far the most popular nonuniform sampling strategy is the log-polar transformation. This mapping was proposed by Schwartz [107] as transformation of the retina to the striate cortex. If u and v are retinal Cartesian coordinates then r and s are the cortical coordinates defined by the following log-polar transformation:

$$r = \log\left(u^2 + v^2\right) \tag{2.1}$$

$$s = \arctan\left(\frac{v}{u}\right). \tag{2.2}$$

Fig. 2.1 illustrates the retinal mapping. Notice how the resolution decreases with the distance form the periphery. Fig. 2.1 also shows the result of mapping the retinal coordinates into the cortical coordinates. The circles are mapped to straight lines. Some benefits of such mapping are apparent. The decreasing resolution toward the periphery provides a wide field-of-view without the attendant explosion in the number of pixels that would occur for uniform sampling. Other advantages such as its aid in rotational and scale invariance are not as apparent. The log-polar transformation has shown success in binocular vision related tasks such as vergence [74], time-to-impact analysis [114], and depth recovery of a scene [23, 103].

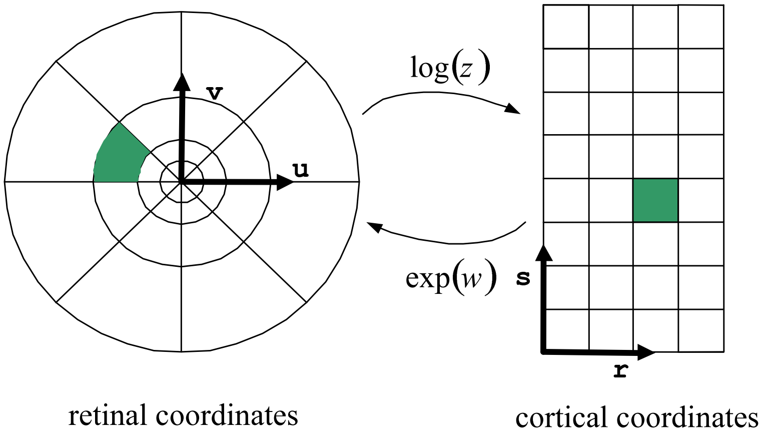


Figure 2.1: Retinal and cortical mappings for log-polar transformation.

In [72], Li and Tong use the Reciprocal Wedge Transform (RWT) as part of an active stereo system. The RWT given by

$$\eta = \frac{1}{x} \tag{2.3}$$

$$\xi = \frac{y}{x} \tag{2.4}$$

simulates the projection onto a nonfrontal image plane, making it advantageous for stereo vision. It can reduce the amount of information by up to 95%. Unlike the log-polar transformation where lines are transformed into log-sine curves, the RWT preserves linear structures. In [72], demonstrated its use in vergence and sparse registration. Unfortunately, the RWT offers no foveal region, putting it somewhat at odds with the active vision paradigm. In general, this transformation has received little attention.

In [110, 136], Aggarwal *et al.* used a fish-eye lense to expand the field-of-view of a normal camera for use on a mobile robot. Since the distorted image was nonlinearly remapped back onto a uniformly sampled image, this application is not a true example of space-variant sampling.

In these above instances the sampling strategies were introduced elsewhere and later applied to areas of stereo vision. Only rarely has a foveation strategy been specifically tailored for a stereo vision task. Both Basu and Elnagar derived optimal sampling schemes with respect to the error discretization of depth measurements. Basu first examined optimal sampling for fronto-parallel stereo cameras [9]. He assumed independent uniform sampling in both x and

y. His goal was to determine the optimal spacing e_x and e_y to minimize 3D reconstruction error subject to the constraint of fixed resolution R:

$$\frac{1}{e_x}\frac{1}{e_y} = R. \tag{2.5}$$

To simplify the problem he chose to optimize only the vertical 3D error; though, he was able to show that errors in the other two dimensions were reduced when compared to uniform sampling (for uniform sampling $e_x = e_y$). The equations describing the optimal solutions for e_x and e_y were polynomials, and a closed form solution was possible.

In [10], Basu next considered error discretization for vergent geometries. Though he still assumed uniform sampling in the y dimension, he proposed using nonuniformly sampling strategies in x. In the x dimension he first considered nonuniform sampling of the form

$$e_x\left(i\right) = E_{min} + \nu i. \tag{2.6}$$

He also examined exponentially increasing sampling:

$$e_x\left(i\right) = E_{min} e^{\gamma i}. \tag{2.7}$$

Again, to make the problem more tractable, Basu chose to optimize only the vertical error. The optimization equations were nonlinear and were solved numerically. Both sampling schemes demonstrated improvement over uniform sampling in all three dimensions.

In [42], Elnagar improved the results obtained by Basu in [9] for fronto-parallel cameras. He was able to solve for the 3D error by optimizing both e_x

and e_y simultaneously. Basu only optimized the vertical error with respect to e_x. Elganar was also able to dispense with all approximations previously used. The optimization equations formed fourth order polynomials that he was able to solve using Mathematica.

Perhaps the most significant use of foveation to specifically address a computer vision task was employed by Klarquist and Bovik [70]. They designed an active vision system called FOVEA. FOVEA was able to reconstruct 3D scenes by dynamically reacting to the specific environment. To simplify the registration process Klarquist and Bovik designed a foveated tessellation that actively adapted to fit the current scene. The tessellation reduced the resolution in the periphery to allow non-horizontal epipolar polar lines to remain in registration. The scheme entailed reducing the resolution toward the periphery by a factor of two each time the epipolar lines diverged by a vertical disparity greater than one-half of a pixel at the current resolution. With this requirement FOVEA could use any off-the-shelf stereo correspondence algorithm that assumed horizontal epipolar lines.

Chapter 3

Local Phase

3.1 Introduction

As an alternative to typical feature based algorithms, phase-based techniques, such as phase-differencing and phase-correlation, have proven to be a powerful means for stereo image correspondence [15,23–26,61,64,99,115,131, 133]. These techniques can produce disparity maps to sub-pixel accuracy, without requiring explicit sub-pixel reconstruction. Since relevant phase information is available at any point in the image, not only at specific tokens as in feature based methods [15,39], phase-based methods yield dense disparity maps. Additionally, phase is relatively insensitive to typical inter-image differences such as lighting, shadows, and noise [35].

Phase-differencing techniques suffer from two significant impediments: phase nonlinearities and neglect of multiscale information. These two concerns are the topics of the following subsections.

3.1.1 Phase Nonlinearities Due to Localization

The premise of phase differencing methods originates in the Fourier Shift Theorem, which states that the Fourier Transforms of two signals related by a global shift are themselves identical up to a phase difference equal to the frequency times the shift. This linear relationship makes it possible to immediately recover this shift or disparity at any frequency (that is present in the signal) by simply dividing the phase difference by the frequency. Unfortunately, stereo images are not related by a global shift, but instead, are better modeled by many local shifts. Consequently, the generation of accurate disparity maps requires a spatially localized estimate of the phase. Such localization is often accomplished by windowing the complex exponential function. The introduction of this windowing function disrupts the simple, linear relation between phase and disparity. In order to successfully apply phase differencing to a pair of stereo images it is necessary to completely understand these disruptive effects. Perhaps the most comprehensive paper in this direction is [48]. Here Fleet *et. al.* models images as one-dimensional Gaussian white noise processes and derives bounds for the expected mean phase difference and absolute mean phase difference under the operations of translation and dilation. In [19], Cai considers the instability of localized phase from the viewpoint that registration by phase differencing can be considered as a type of Newton iteration. He points out that the instability of phase measurements in certain regions violates the criteria necessary for convergence.

A key to successfully applying phase differencing stereo reconstruction algorithms lies in the identification of regions of phase instability, i.e. the portions of signals that possess specific properties that invalidate the use of phase for disparity estimation. Several papers address this issue [48,61,63,90], focusing primarily on the same two means for instability detection. Though these techniques have been shown to be empirically successful, their foundations are somewhat heuristic and their efficacy has never been fully examined analytically.

One intent of this work is to supplement the existing literature by filling in some important gaps in the current understanding of localized phase for disparity estimation. Modeling the images as Gaussian white noise processes, we analytically quantify the behavior of the the first and second derivatives of local phase. The first derivative describes the linear relation between phase and disparity. The second derivative provides valuable information about the nonlinearity of localized phase. We demonstrate that the current methods for identifying regions of instability are simply approximations of the second derivative. Finally, we introduce a new method for detecting phase instability that fully incorporates the information contained in the second derivative.

3.1.2 Neglect of Multiscale Information

Phase-differencing algorithms require coarse to fine processing schemes to allow larger scales to bring the images into near registration before proceeding to higher scales. At each scale disparity information is procured from a small range of frequencies. Images with relatively little energy in these frequency bands will tend to produce erroneous estimates that will further contaminate higher scales as the iteration proceeds. To mitigate this problem, alternative means for using phase information have been suggested. A typical approach is to apply a local Fourier Transformation [47,129] or wavelet decomposition [73, 133] and then globally search between images for locations that minimize the difference in their coefficients. This difference is predominately a function of the local phase. These methods are time consuming, disregarding the linearity of phase that allows phase differencing to rapidly convergence to high resolution solutions. Furthermore, their resolution is inherently limited by the sampling of the transformations, and consequently require ancillary methods for sub-pixel estimation. Other approaches use phase differencing to realize disparity estimates at multiple frequencies then heuristically combine them at the end [87,132]. Phase information has also been combined over multiple frequencies using probabilistic schemes [115], spatial representations [41], and image derivatives [88].

The second intent of this dissertation is to introduce a novel algorithm for combining multiscale disparity information in a theoretically sound approach that retains the fast, high resolution convergence properties of phase differencing. Developing this approach, we demonstrate the intimate relation between phase differencing and local correlation.

3.2 Phase-Differencing for Stereo Correspondence

In this section we first present the phase differencing algorithm and review its underlying theory. Next we discuss conditions under which phase differencing is inappropriate. We then offer a new paradigm for identifying these conditions and demonstrate how previous methods of identification that are prominent in the literature are better understood in light of this paradigm.

3.2.1 Review

Given a pair of stereo images $i_1(x)$ and $i_2(x)$, the goal of a stereo correspondence algorithm is to match each point in $i_1(x)$ with its corresponding point in $i_2(x)$. If the point x_o in $i_1(x)$ corresponds to the point $x_o - \delta$ in $i_2(x)$ then δ is said to be the disparity relating $i_1(x)$ to $i_2(x)$ at x_o.

The methodology for recovering disparity by phase-difference is simple. Each image is convolved with a complex kernel $h(x)$ whose real and imaginary parts are Hilbert Transforms of one another (i.e. quadrature filters). For our purposes we assume the filter is formed by windowing the complex exponential with a real, symmetric function as follows: $h(x) = g(x)\, e^{j\omega_o x}$. The windowing function $g(x)$ must be chosen appropriately such that real and imaginary parts of $h(x)$ remain Hilbert Transforms of each other. If $O_1(x)$ and $O_2(x)$ represent the convolutions of $i_1(x)$ and $i_2(x)$ with $h(x)$, then the disparity is estimated with a single calculation: $\delta(x) = \Delta\theta(x) / \omega_o$, where $\Delta\theta(x) = \arg[O_2(x)] - \arg[O_1(x)]$ is the difference in phase between the two responses.

For simplicity, let us only consider the responses at the point $x_o = 0$. This allows us to replace the convolution with the inner product. Accordingly, we now have $O_j =< i_j(x), g(x)\, e^{j\omega_o x} >$, where $< f_1, f_2 >= \int f_1^* f_2 dx$. The equation to recover disparity becomes

$$\delta = \frac{\Delta\theta}{\omega_o}. \tag{3.1}$$

To better understand the principles governing phase differencing consider the simple case where the windowing function $g(x) = 1$ and the images $i_1(x)$ and $i_2(x)$ are related by a single, global translation, i.e. $i_2(x) = i_1(x - \delta)$. Now we have $O_1 = I(-\omega_o)$ and $O_2 = I(-\omega_o) e^{j\omega_o \delta}$, where $I(w)$ is the Fourier Transform of $i(x)$. Furthermore, $\Delta\theta = \arg[O_2] - \arg[O_1] = \omega_o \delta$, and δ can be recovered using (3.1). For complex exponentials the relationship between phase and translation is linear with the relational constant ω_o.

Since actual stereo images are not generally related by a single global shift, but instead, by many local shifts (and other deformations), the window $g(x) = 1$ with its infinite spatial extent is not very useful. The apparent solution would seem to be the localization of the window. Unfortunately, there is no guarantee that disparity will remain a linear function of phase. That is, equation (3.1) may no longer hold. Fortunately, it turns out that phase is linear except in limited regions of phase instability. The identification of these regions of nonlinear phase is the topic of the next section.

3.2.2 Regions of Phase Nonlinearity

In this section we discuss the spatial localization of the window $g(x)$ and its effects on the linearity of phase. We begin by assuming that within the window's spatial extent, defined as the region containing nearly all its energy, the functions $i_1(x)$ and $i_2(x)$ are modeled reasonably well by a single translation δ, i.e. $i_2(x) = i_1(x - \delta)$ within $g(x)$. Let us now express O as a function of δ: $O(\delta) =< i(x - \delta), g(x) e^{j\omega_o x} >$, $O_1 = O(0)$ and $O_2 = O(\delta)$.

Henceforth, we will refer only to i and O and not i_1, i_2, O_1, and O_2. Note that a translation of the image is equivalent to an opposite translation of the filter since $< i\left(x-\delta\right), g\left(x\right)e^{j\omega_o x} >=< i\left(x\right), g\left(x+\delta\right)e^{j\omega_o(x+\delta)} >$. Since local phase is analytic everywhere except at the zeros, we can use this relation to expand $\Delta\theta$ into a Taylor Series:

$$\Delta\theta = \arg\left[O\left(\delta\right)\right] - \arg\left[O\left(0\right)\right] = \tilde{\omega}\delta + 1/2\tilde{\omega}'\delta^2 + \bigcirc\left(\delta^3\right), \qquad (3.2)$$

where the instantaneous frequency $\tilde{\omega}$ is the constant of linearity and its derivative $\tilde{\omega}'$ indicates the degree of nonlinearity. If the phase difference is dominated by the linear term $\tilde{\omega}$, the disparity can be recovered in a single step by using this modified version of (3.1):

$$\delta = \frac{\Delta\theta}{\tilde{\omega}}. \qquad (3.3)$$

Two situations arise that may prevent the accurate recovery of phase from (3.3). First the phase difference may not be linearly related to the disparity, i.e. the higher order terms of (3.2) may dominate the linear term. Second the relationship may be linear, but $\tilde{\omega}$ may be so large as to preclude the detection of disparities beyond a maximum value due to phase wrapping. Phase wrapping forces $|\Delta\theta| < \pi$, and consequently, $|\delta| < \pi/\tilde{\omega}$ from (3.3).

The literature promotes two constraints to avoid areas of phase instability (or detect them by identifying regions where the constraints are violated) [61, 99]. The first constraint requires that $\tilde{\omega}$ not deviate too far from the center frequency ω_o:

$$|\tilde{\omega} - \omega_o| < \rho_1. \qquad (3.4)$$

This requirement is often given in the context of detecting phase instability due to scaling deformations between views, but we will see it is also very valuable for nonlinearities due to translation. The second requires that

$$\frac{\left| |O|_x \right|}{|O|} < \rho_2, \tag{3.5}$$

where $|O| = \sqrt{\mathrm{Re}\,[s]^2 + \mathrm{Im}\,[s]^2}$ and $|O|_x$ is the partial derivative of $|O|$ with respect to x. This requirement is a consequence of the instability of phase near singularities, i.e. points where $|O| = 0$ and phase is undefined. The first order estimate of the distance to such a point is $|O| \,/\, \left||O|_x\right|$. Inverting this ratio we get (3.5).

For simplicity, sometimes these two constraints are combined into one [48]

$$\sqrt{(\tilde{\omega} - \omega_o)^2 + \left(\frac{|O|_x}{|O|}\right)^2} < \rho_3. \tag{3.6}$$

We will later show that though this combination seems to be a result of convenience, it is actually the more appropriate condition. This claim was indirectly asserted by Cai [19] when he demonstrated that the region of convergence of a Newton iteration using (3.3) is the disk bounded by (3.6).

To simplify further discussions we present the following definitions:

$$\begin{aligned} a + jb &= < i\,(x), g\,(x)\, e^{j\omega_o(x)} > \\ c + jd &= < i\,(x), g'\,(x)\, e^{j\omega_o(x)} > \\ e + jf &= < i\,(x), g''\,(x)\, e^{j\omega_o(x)} >, \end{aligned} \tag{3.7}$$

where, for example, $a = \int i\,(x) \cos{(\omega_o x)} g\,(x)\, dx$. Using these definitions we can express θ, $\tilde{\omega}$, and $\tilde{\omega}'$ as follows:

$$\theta = \tan^{-1}\left(\frac{b}{a}\right) \tag{3.8}$$

$$\tilde{\omega} = \frac{\partial}{\partial x}\left[\tan^{-1}\left(\frac{b}{a}\right)\right] = \omega_o + \frac{ad - bc}{a^2 + b^2} = \omega_o + \xi \tag{3.9}$$

$$\begin{aligned} \tilde{\omega}' &= \frac{\partial}{\partial x}\left[\omega_o + \frac{ad - bc}{a^2 + b^2}\right] = \frac{af - be}{a^2 + b^2} - 2\frac{ad - bc}{a^2 + b^2}\frac{ac + bd}{a^2 + b^2} \\ &= \tau - 2\xi\chi = \tau - 2\nu. \end{aligned} \tag{3.10}$$

The constraints presented in (3.4), (3.5), and (3.6) can also be reprised using (3.7):

$$|\tilde{\omega} - \omega_o| = |\xi| < \rho_1 \tag{3.11}$$

$$\frac{||O|_x|}{|O|} = \left|\frac{ac + bd}{a^2 + b^2}\right| = |\chi| < \rho_2. \tag{3.12}$$

$$\sqrt{(\tilde{\omega} - \omega_o)^2 + \left(\frac{|O|_x}{|O|}\right)^2} = \sqrt{\xi^2 + \chi^2} < \rho_3. \tag{3.13}$$

Interestingly, the linear term $\tilde{\omega}$ in (3.9) consists of ω_o, corresponding to the center frequency of the complex exponential, and the additional term ξ induced by the windowing function. We should also note that ξ is precisely the term constrained in (3.11). More importantly, $\tilde{\omega}'$, the term in (3.2) that most directly indicates the degree of nonlinearity, contains the two most prominently advocated constraints in detecting regions of phase instability, ξ and χ. By limiting the values of these variables, these conditions have in fact constrained the possible values of $\tilde{\omega}'$. We will delve into these issues at greater depth in the following sections.

3.3 Gaussian Random Images

In this section we quantify the effect that the spatial localization of $g(x)$ has on the linear relation between phase and disparity. In order to produce analytic results the image $i(x)$ will be modeled as the stationary Gaussian white noise process $\boldsymbol{i}(x)$. The nice properties of Gaussian random variables (GRVs) will allow us to analytically derive simple, closed-form results for many important properties of localized phase. We should note that white noise can not model all aspects of real images. Phenomena such as occlusions, foreshortening, and shadowing are disregarded. Furthermore, white noise is spread spectrum, while real images tend to concentrate their energy near DC.

Since $\boldsymbol{i}(x)$ is a Gaussian random process, the integrals in (3.7) become GRVs. In the following sub-sections we first determine the means, variances, and covariances of these variables. We then derive relevant distributions and moments that will help us better understand $\tilde{\omega}$ and $\tilde{\omega}'$ and their importance to linear phase. Finally, we quantify the effects of the constraints in (3.11), (3.12), and (3.13) in relation to these distributions and moments.

3.3.1 Means, Variances, and Covariances

In this section we explore the mean, variances, and covariances of the the random variables that result from (3.7) when $\boldsymbol{i}(x)$ is a white noise process. The random variables $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$, $\boldsymbol{e}$, and $\boldsymbol{f}$ are all zero mean due to the zero-mean ergodicity of $\boldsymbol{i}(x)$. Since the kernel $g(x)\, e^{j\omega_o x}$ is defined to be a quadrature filter (its real and imaginary components are Hilbert transforms of each other) the variances become

$$\sigma_a^2 = \mathrm{E}[\boldsymbol{a^2}] = \int [\cos(\omega_o x) g(x)]^2 \, dx = \int [\sin(\omega_o x) g(x)]^2 \, dx = \sigma_b^2 \quad (3.14)$$

$$\sigma_b^2 = \mathrm{E}[\boldsymbol{c^2}] = \int [\cos(\omega_o x) g'(x)]^2 \, dx = \int [\sin(\omega_o x) g'(x)]^2 \, dx = \sigma_c^2 \quad (3.15)$$

$$\sigma_e^2 = \mathrm{E}[\boldsymbol{e^2}] = \int [\cos(\omega_o x) g''(x)]^2 \, dx = \int [\sin(\omega_o x) g''(x)]^2 \, dx = \sigma_f^2 \quad (3.16)$$

Furthermore, the orthogonality between the real and imaginary parts of quadrature filters necessitates that $\mathrm{E}[\boldsymbol{ab}] = 0$, $\mathrm{E}[\boldsymbol{cd}] = 0$, and $\mathrm{E}[\boldsymbol{ef}] = 0$. In fact, the variables $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ are all uncorrelated (and hence independent). As proof, consider the following:

$$\begin{aligned}
\mathrm{E}[\boldsymbol{ad}] &= \iint \delta(x-y) \cos(\omega_o x) \sin(\omega_o y) g(x) g'(y) \; dx \, dy \\
&= \frac{\omega_o}{2} \int \left[\sin^2(\omega_o x) - \cos^2(\omega_o x)\right] g^2(x) \; dx \\
&= \frac{\omega_o}{2} \left(\mathrm{E}[\boldsymbol{b}^2] - \mathrm{E}[\boldsymbol{a}^2]\right) = 0. \qquad (3.17)
\end{aligned}$$

Verifying that the remaining combinations are also zero requires either following the previous derivation or noticing that the integrands are odd functions.

Some cross-correlations involving the random variables $\boldsymbol{e}$ and $\boldsymbol{f}$ are generally nonzero.

$$\begin{aligned}
\mathrm{E}[\boldsymbol{ae}] &= \iint \delta(x-y) \cos(\omega_o x) \cos(\omega_o y) g''(x) g(y) \; dx \, dy \\
&= \int g''(x) g(x) \cos^2(\omega_o x) \, dx \\
&= 2\omega_o \int g(x) g' \cos(\omega_o x) \sin(\omega_o x) \, dx - \int \left[g'(x)\right]^2 cos^2(\omega_o x) \, dx \\
&= 2\omega_o \mathrm{E}[\boldsymbol{ad}] - \mathrm{E}\left[\boldsymbol{d}^2\right] = -\sigma_c^2 \qquad (3.18)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[\boldsymbol{bf}] &= \int\int \delta(x-y)\sin(\omega_o x)\sin(\omega_o y)g''(x)g(y)\;dx\,dy \\
&= \int g''(x)g(x)\sin^2(\omega_o x)\,dx \\
&= -2\omega_o \int g(x)g'\cos(\omega_o x)\sin(\omega_o x)\,dx - \int \left[g'(x)\right]^2 sin^2(\omega_o x)\,dx \\
&= -2\omega_o \mathrm{E}\left[\boldsymbol{ad}\right] - \mathrm{E}\left[\boldsymbol{c}^2\right] = -\sigma_c^2 \qquad (3.19)
\end{aligned}$$

Equations (3.18) and (3.19) follow from integration by parts and then substitution using (3.17). The remaining cross-correlations are zero. Their integrands either form odd functions or follow the derivation below.

$$\begin{aligned}
\mathrm{E}[\boldsymbol{cf}] &= \iint \delta(x-y)\sin(\omega_o x)\sin(\omega_o y)g''(x)g'(y)dxdy \\
&= \frac{\omega_o}{2}\int g''(x)g'(x)\cos(\omega_o x)\sin(\omega_o x)dx \\
&= \int \left[sin^2(\omega_o x) - \cos^2(\omega_o x)\right]\left[g'(x)\right]^2 dx \\
&= \frac{\omega_o}{2}\left(\mathrm{E}\left[\boldsymbol{c}^2\right] - \mathrm{E}\left[\boldsymbol{d}^2\right]\right) = 0 \qquad (3.20)
\end{aligned}$$

In summary, the covariance matrix K relating $\boldsymbol{a}$, $\boldsymbol{c}$, and $\boldsymbol{e}$ is identical to the covariance matrix relating $\boldsymbol{b}$, $\boldsymbol{d}$, and $\boldsymbol{f}$:

$$K = \begin{bmatrix} \sigma_a^2 & 0 & -\sigma_c^2 \\ 0 & \sigma_c^2 & 0 \\ -\sigma_c^2 & 0 & \sigma_e^2 \end{bmatrix} = \begin{bmatrix} \sigma_b^2 & 0 & -\sigma_d^2 \\ 0 & \sigma_d^2 & 0 \\ -\sigma_d^2 & 0 & \sigma_f^2 \end{bmatrix} \qquad (3.21)$$

The random variables $\boldsymbol{a}$, $\boldsymbol{c}$, and $\boldsymbol{e}$ are independent of $\boldsymbol{b}$, $\boldsymbol{d}$, and $\boldsymbol{f}$.

3.3.2 Instantaneous Frequency

The instantaneous frequency $\tilde{\omega}$ is expressed as the sum of two terms in (3.9). The first term ω_o represents the center frequency of the bandpass filter $g(x)e^{j\omega_o x}$ and corresponds to the constant change in phase caused by a global shift in $i\left(x\right)$ when the windowing function $g(x)$ is constant and of infinite support. Localizing $e^{j\omega_o x}$ with $g(x)$ induces the second term ξ. The random variable $\boldsymbol{\xi}$ corresponding to this localization term is a function of the four independent, zero-mean, Gaussian random variables $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$. Its distribution (see Appendix 1.1.1) is

$$f\left(\xi\right)=\frac{\sigma_c^2/\sigma_a^2}{2\left(\sigma_c^2/\sigma_a^2+\xi^2\right)^{\frac{3}{2}}}. \tag{3.22}$$

The random variable $\boldsymbol{\xi}$ is zero mean with infinite variance. The absolute deviation exists:

$$\mathrm{E}\left[\left|\boldsymbol{\xi}\right|\right]=\sqrt{\sigma_c^2/\sigma_a^2}. \tag{3.23}$$

3.3.3 The Derivative of the Instantaneous Frequency

Analysis of the derivative of the instantaneous frequency $\tilde{\omega}'$ presents a greater challenge; determining its probability density function in a closed form appears to be an intractable problem. We instead independently examine its two constitute components, τ and ν as delineated in (3.10). The form of $\boldsymbol{\tau}$ is very similar to that of $\boldsymbol{\xi}$. The difference lies in the correlation among the random variables that comprise it. The random variables $\boldsymbol{a}$ and $\boldsymbol{f}$ are correlated, as are $\boldsymbol{b}$ and $\boldsymbol{e}$. The resultant pdf of $\boldsymbol{\tau}$ (see Appendix 1.1.1) is

$$f\left(\tau\right)=\frac{|K|/\left(\sigma_a^4\sigma_c^2\right)}{2(|K|/\left(\sigma_a^4\sigma_c^2\right)+\tau^2)^{\frac{3}{2}}}, \tag{3.24}$$

where K is given in (3.21). Similar to the localization term $\boldsymbol{\xi}$, $\boldsymbol{\tau}$ is also zero mean with infinite variance. Its absolution deviation is finite with value

$$\mathrm{E}\left[|\boldsymbol{\tau}|\right] = \sqrt{|K|/\left(\sigma_a^4\sigma_c^2\right)}. \tag{3.25}$$

The second term $\boldsymbol{\nu}$ is the product of $\boldsymbol{\xi}$ and $\boldsymbol{\chi}$, also shown in (3.10). The RV $\boldsymbol{\xi}$ is the localization term of $\tilde{\boldsymbol{\omega}}$ discussed in Section 3.3.2. Equation (3.10) indicates that both $\boldsymbol{\xi}$ and $\boldsymbol{\chi}$ are slightly different functions of the same RVs: $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$. To better elucidate their interrelation we convert them into polar form, setting $\boldsymbol{a} = \boldsymbol{r_1}\cos(\boldsymbol{\phi_1})$, $\boldsymbol{b} = \boldsymbol{r_1}\sin(\boldsymbol{\phi_1})$, $\boldsymbol{c} = \boldsymbol{r_2}\cos(\boldsymbol{\phi_2})$, and $\boldsymbol{d} = \boldsymbol{r_2}\sin(\boldsymbol{\phi_2})$. The random variables $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ have Rayleigh distributions while $\boldsymbol{\phi_1}$ and $\boldsymbol{\phi_2}$ are distributed uniformly over the interval $[-\pi : \pi]$ [96]. Substituting these values we find $\boldsymbol{\xi} = \boldsymbol{r_2}/\boldsymbol{r_1}\sin\left(\boldsymbol{\phi_2} - \boldsymbol{\phi_1}\right)$ and $\boldsymbol{\chi} = \boldsymbol{r_2}/\boldsymbol{r_1}\cos\left(\boldsymbol{\phi_2} - \boldsymbol{\phi_1}\right)$. We can simplify further by letting $\boldsymbol{\phi} = \boldsymbol{\phi_2} - \boldsymbol{\phi_1}$, where $\boldsymbol{\phi}$ is uniformly distributed over $[-\pi : \pi]$. The final forms of $\boldsymbol{\xi}$, $\boldsymbol{\chi}$, and $\boldsymbol{\nu}$ are

$$\boldsymbol{\xi} = \frac{\boldsymbol{r_2}}{\boldsymbol{r_1}}\sin\left(\boldsymbol{\phi}\right) \tag{3.26}$$

$$\boldsymbol{\chi} = \frac{\boldsymbol{r_2}}{\boldsymbol{r_1}}\cos\left(\boldsymbol{\phi}\right) \tag{3.27}$$

$$\boldsymbol{\nu} = \boldsymbol{\xi}\boldsymbol{\chi} = \left(\frac{\boldsymbol{r_2}}{\boldsymbol{r_1}}\right)^2\cos\left(\boldsymbol{\phi}\right)\sin\left(\boldsymbol{\phi}\right). \tag{3.28}$$

The RV $\boldsymbol{\nu}$ has zero mean, infinite variance, and an infinite absolute deviation (see Appendix 1.2).

3.3.4 Effects of Constraints

In the previous subsections we found the distribution for the localization term $\boldsymbol{\xi}$ of $\tilde{\boldsymbol{\omega}}$ and the distributions for the components $\boldsymbol{\tau}$ and $\boldsymbol{\nu}$ that comprise $\tilde{\boldsymbol{\omega}}'$. All three of these pdfs have heavy tails. In fact, all have infinite variances. The distribution for $\tilde{\boldsymbol{\omega}}'$ does not even have a finite absolute deviation. This is not unexpected. Equations (3.9) and (3.10) both indicate that the instantaneous frequency $\tilde{\omega}$ and the derivative of instantaneous frequency $\tilde{\omega}'$ are comprised of components that are the ratios of sums. Since the summations in their denominators are zero mean for *any* random, zero mean input signal $i\,(x)$, we would expect these ratios to frequently produce very high values. High values result in erroneous disparity estimates for one of two reasons. First, extreme values of $\tilde{\omega}$ produce incorrect estimates due to phase wrapping. Second, large values of $\tilde{\omega}'$ indicate that the constant of linearity is rapidly changing. Consequently, the assumption of linearity upon which phase differencing relies becomes invalid.

It is important reiterate that though we have used GRVs to analytically quantify the extent to which $\tilde{\boldsymbol{\omega}}$ and $\tilde{\boldsymbol{\omega}}'$ produce extreme values, the cause of these extreme values lies not with the assumption of normality, but instead, with the ratios inherent in their specific representations. This fact is further substantiated in the later sections where the distributions of actual images are considered.

In Section 3.2.2 we delineated the constraints commonly used to identify regions of phase instability. We now examine the effects of these constraints on the random variables $\boldsymbol{\xi}$, $\boldsymbol{\tau}$ and $\boldsymbol{\nu}$. The first constraint given in (3.11) requires that the localization term $\boldsymbol{\xi}$ fall within a certain range. This will, in general, reduce the spread of the RV's $\boldsymbol{\xi}$, $\boldsymbol{\tau}$, and $\boldsymbol{\nu}$ (see Appendix 1.3):

$$\mathrm{E}\left[|\boldsymbol{\xi}| \,\big|\, |\xi| < \rho_1\right] = \frac{1}{\rho_1}\left(\frac{\sigma_c}{\sigma_a}\left[\frac{\sigma_c^2}{\sigma_a^2} + \rho_1^2\right]^{\frac{1}{2}} - \frac{\sigma_c^2}{\sigma_a^2}\right) \tag{3.29}$$

$$\mathrm{E}\left[|\boldsymbol{\tau}| \,\big|\, |\xi| < \rho_1\right] = \frac{2}{\pi\rho_1}\left(\frac{|K|}{\sigma_a^4\sigma_c^2}\right)^{\frac{1}{2}}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_1^2\right)^{\frac{1}{2}}\tan^{-1}\left(\frac{\sigma_a}{\sigma_c}\rho_1\right) \tag{3.30}$$

$$\mathrm{E}\left[|\boldsymbol{\nu}| \,\big|\, |\xi| < \rho_1\right] = \frac{1}{\pi\rho_1}\frac{\sigma_c^2}{\sigma_a^2}\left(\frac{\sigma_a^2}{\sigma_c^2} + \rho_1^2\right)^{\frac{1}{2}}\ln\left(1 + \frac{\sigma_c^2}{\sigma_a^2}\rho_1^2\right). \tag{3.31}$$

In fact, the variances of all three RV's are now finite, whereas previously their variances were infinite as was the absolution deviation of $\boldsymbol{\nu}$.

Before considering the second constraint given in (3.12), we will reexamine both constraints in light of the polar representations of $\boldsymbol{\xi}$ and $\boldsymbol{\chi}$ shown in (3.26) and (3.27). In polar coordinates the constraints become:

$$|\xi| = \left|\frac{r_2}{r_1}\sin(\phi)\right| < \rho_1 \tag{3.32}$$

$$|\chi| = \left|\frac{r_2}{r_1}\cos(\phi)\right| < \rho_2. \tag{3.33}$$

Using these constraints in the fashion suggested by (3.11) and (3.12) amounts to forming a rectangular decision region in (ξ, χ) space. Observing the polar forms of the constraints and remembering that $\boldsymbol{\phi}$ is uniformly distributed suggests that a circular decision region would be more appropriate:

$$\sqrt{\chi^2 + \xi^2} = \sqrt{\left(\frac{r_2}{r_1}\right)^2\left(\cos^2(\phi) + \sin^2(\phi)\right)} = \frac{r_2}{r_1} < \rho_3. \tag{3.34}$$

Incorporating this constraint we find that (see Appendix 1.3)

$$\mathrm{E}\Big[|\boldsymbol{\xi}| \,\big|\, \sqrt{\chi^2 + \xi^2} < \rho_3\Big] = \frac{2}{\pi\rho_3}\frac{\sigma_c^2}{\sigma_a^2}\left(\frac{1}{\rho}\frac{\sigma_a}{\sigma_c}\tan^{-1}\left[\frac{\sigma_a}{\sigma_c}\rho_3\right]\left[\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right] - 1\right) \tag{3.35}$$

$$\mathrm{E}\Big[|\boldsymbol{\tau}|\,\big|\sqrt{\chi^2+\xi^2}<\rho_3\Big] = \frac{1}{\rho_3^2}\frac{\sigma_c}{\sigma_a}\left[\left(\frac{|K|}{\sigma_a^4\sigma_c^2}\right)\left(\frac{\sigma_c^2}{\sigma_a^2}+\rho_3^2\right)\left(\frac{\sigma_a}{\sigma_b}\left[\frac{\sigma_c^2}{\sigma_a^2}+\rho_3^2\right]-1\right)\right]^{\frac{1}{2}} \tag{3.36}$$

$$\mathrm{E}\Big[|\boldsymbol{\nu}|\,\big|\sqrt{\chi^2+\xi^2}<\rho_3\Big] = \frac{1}{\pi}\frac{\sigma_c^2}{\sigma_a^2}\left(\frac{\sigma_c^2/\sigma_a^2+\rho_3^2}{\rho_3^2}\left[\ln\left(\frac{\sigma_c^2/\sigma_a^2+\rho_3^2}{\sigma_c^2/\sigma_a^2}\right)\right]-1\right) \tag{3.37}$$

Using the triangle inequality, we can now establish a bound on the expected nonlinearity $\mathrm{E}[|\tilde{\boldsymbol{\omega}}'|]$ in regions satisfying (3.13):

$$\begin{aligned}\mathrm{E}\left[|\tilde{\boldsymbol{\omega}}'|\,\big|\sqrt{\chi^2+\xi^2}<\rho_3\right] =& \mathrm{E}\left[|\boldsymbol{\tau}-2\boldsymbol{\nu}|\,\big|\sqrt{\chi^2+\xi^2}<\rho_3\right] \\ \leq& \mathrm{E}\left[|\boldsymbol{\tau}|\,\big|\sqrt{\chi^2+\xi^2}<\rho_3\right] + \\ & 2\mathrm{E}\left[|\boldsymbol{\nu}|\,\big|\sqrt{\chi^2+\xi^2}<\rho_3\right]\end{aligned} \tag{3.38}$$

We have now demonstrated that the constraint in (3.13) serves to limit the second derivative of phase $\tilde{\omega}'$. In fact, the second term ν can be completed controlled by this constraint. Unfortunately, though the first term τ also tends to be reduced, it can still obtain arbitrarily large values. To further restrict this term, and consequently $\tilde{\omega}'$, we introduce the additional constraint

$$|\tau| < \rho_4. \tag{3.39}$$

The rationale and appropriateness of this condition will become more apparent in later sections when actual kernels are evaluated.

It is important to put the results of this section into context. We previously mentioned that extreme values of $\tilde{\omega}$ and $\tilde{\omega}'$ can invalidate the use of phase differencing for disparity estimation. This section demonstrated that by restricting the use of phase differencing to regions satisfying the above constraints, the frequency of the extreme value can be significantly reduced, improving the accuracy of disparity estimates.

3.4 Gabor Functions

In this section we use the previous results to analyze a specific kernel, the Gabor function. For a Gabor kernel $h(x)$ the windowing function $g(x)$ in (3.7) is the Gaussian. The Gaussian is often the window of choice because of its joint optimality with respect to support in both time and space [52]. In addition to stereo registration [23, 24, 26], the Gabor function has found many useful applications in areas such as texture analysis [30, 31]. Furthermore, the Gabor function is a common model of the simple receptive fields in the visual cortex [6, 29, 65]. Normalizing $h(x)$ to have unit energy yields

$$h(x) = \frac{1}{\left(\sqrt{\pi}\sigma_g\right)^{\frac{1}{2}}} e^{-\frac{x^2}{2\sigma_g^2}} e^{j\omega_o x}. \tag{3.40}$$

The Fourier Transform of the Gabor kernel is a Gaussian shifted by the modulation frequency ω_o with standard deviation $\sigma_\omega = 1/\sigma_g$. In order for a Gabor function to adequately approximate a quadrature filter, ω_o must be much greater than σ_ω. To ensure this relationship across different frequencies Gabor functions of constant relative bandwidth β are usually assumed, i.e. $\beta = \log_2\left[\left(\omega_o + \sigma_\omega\right) / \left(\omega_o - \sigma_\omega\right)\right]$. A bandwidth less than or equal to one octave ($\beta \leq 1$) is usually sufficient to well approximate a quadrature pair filter. With $\beta = 1$ the Gaussian envelope in the frequency domain crosses DC at three standard deviations. A perfect quadrature pair filter requires a zero response at DC. For a more complete discussion on the selection of parameters for Gabor functions see [36].

Incorporating $h\left(x\right)$ into (3.7), the covariance matrix in (3.21) becomes

$$K = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{4\sigma_g^2} \\ 0 & \frac{1}{4\sigma_g^2} & 0 \\ -\frac{1}{4\sigma_g^2} & 0 & \frac{3}{8\sigma_g^4} \end{bmatrix}. \tag{3.41}$$

Using these values we can now illustrate the effects of the constraints on the first and second derivatives of phase as formulated in (3.23), (3.25), (3.29), (3.30), (3.31), (3.35), (3.36), and (3.37).

$$\mathrm{E}[|\boldsymbol{\xi}|] = \frac{1}{\sqrt{2}\sigma_g} \tag{3.42}$$

$$\mathrm{E}[|\boldsymbol{\tau}|] = \frac{1}{\sqrt{2}\sigma_g^2} \tag{3.43}$$

$$\mathrm{E}\big[|\boldsymbol{\xi}|\,\big|\,|\boldsymbol{\xi}| < \tilde{\rho}_1\big] = \frac{1}{2\tilde{\rho}_1\sigma_g}\left[\left(1+2\tilde{\rho}_1^2\right)^{\frac{1}{2}} - 1\right] \tag{3.44}$$

$$\mathrm{E}\big[|\boldsymbol{\tau}|\,\big|\,|\boldsymbol{\xi}| < \tilde{\rho}_1\big] = \frac{1}{\pi\tilde{\rho}_1\sigma_g^2}\left(1+2\tilde{\rho}_1^2\right)^{\frac{1}{2}}\tan^{-1}\left(\sqrt{2}\tilde{\rho}_1\right) \tag{3.45}$$

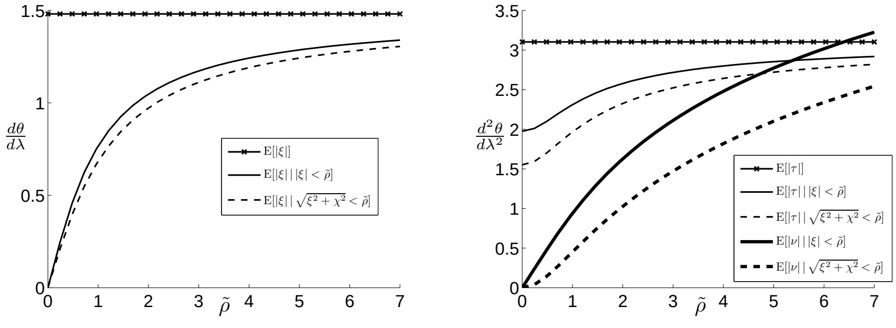
$$\mathrm{E}\big[|\boldsymbol{\nu}|\,\big|\,|\boldsymbol{\xi}| < \tilde{\rho}_1\big] = \frac{1}{2\sqrt{2}\pi\tilde{\rho}_1\sigma_g^2}\left(1+2\tilde{\rho}_1^2\right)^{\frac{1}{2}}\ln\left(1+2\tilde{\rho}_1^2\right) \tag{3.46}$$

$$\mathrm{E}\left[|\boldsymbol{\xi}|\,\Big|\sqrt{\boldsymbol{\xi}^2+\boldsymbol{\chi}^2} < \tilde{\rho}_3\right] = \frac{1}{\sqrt{2}\pi\tilde{\rho}_3{}^2\sigma_g}\left[\left(1+2\tilde{\rho}_3^2\right)\tan^{-1}\left(\sqrt{2}\tilde{\rho}_3\right) - \sqrt{2}\tilde{\rho}_3\right] \tag{3.47}$$

$$\mathrm{E}\left[|\boldsymbol{\tau}|\,\Big|\sqrt{\boldsymbol{\xi}^2+\boldsymbol{\chi}^2} < \tilde{\rho}_3\right] = \frac{1}{2\sqrt{2}\tilde{\rho}_3^2\sigma_g^2}\left(1+2\tilde{\rho}_3^2\right)^{\frac{1}{2}}\left[\left(1+2\tilde{\rho}_3^2\right)^{\frac{1}{2}} - 1\right] \tag{3.48}$$

$$\mathrm{E}\left[|\boldsymbol{\nu}|\,\Big|\sqrt{\boldsymbol{\xi}^2+\boldsymbol{\chi}^2} < \tilde{\rho}_3\right] = \frac{1}{4\pi\tilde{\rho}_3^2\sigma_g^2}\left[\left(1+2\tilde{\rho}_3^2\right)\ln\left(1+2\tilde{\rho}_3^2\right) - 2\tilde{\rho}_3^2\right], \tag{3.49}$$

Fig. 3.1 plots these equations for Gabor kernels with relative bandwidth $\beta = 1$. The X axes indicate the value of the normalized constraint $\tilde{\rho}_i$, where $\rho_i/\sigma_\omega =$



(a) (b)

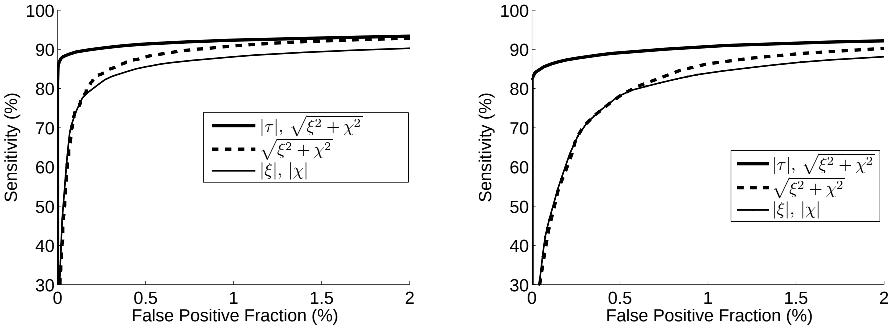
Figure 3.1: $E[|\boldsymbol{\xi}|]$, $E[|\boldsymbol{\tau}|]$, and $E[|\boldsymbol{\nu}|]$ with respect to constraints for Gabor kernels ($\beta = 1$) responding to white noise.

$\tilde{\rho}_i$. Since the Y axes are expressed in terms of wavelength λ_o, all Gabor kernels with the same relative bandwidth produce identical curves. This is a consequence of the following relation: $d\theta/d\lambda = d\theta/dx\ dx/d\lambda = d\theta/dx\ \lambda_o$, where $\lambda_o = 2\pi/\omega_o$.

We next test the ability of the constraints in (3.11), (3.12), (3.13), and (3.39) to identifying regions where phase-based disparity measurements are poor. The constraints can be considered features; and they can be measured at every point we calculate disparity. These features (or some functions of them) can be used to make a decision as to whether we can trust the disparity estimate at that point. Consequently, the points can be separated into two classes: points at which (3.3) produces correct disparity estimates (class $\mathcal{C}_1$) and points at which (3.3) produces incorrect disparity estimates (class $\mathcal{C}_2$). The goal is to use a set of features that will allow us to include the greatest number of points in $\mathcal{C}_1$ while excluding the greatest number of points in $\mathcal{C}_2$.

Prior to evaluating the performance of different sets of features, let us consider the candidate features themselves. In order to detect regions of phase nonlinearity, the literature advocates using the features $|\xi|$ and $|\chi|$ by thresholding them above a certain value. This forms a rectangular decision region in (ξ, χ) space. In Section 3.3.4, we suggested that a circular decision region was more appropriate. Such a region is constructed by thresholding the one dimensional feature $\sqrt{\xi^2 + \chi^2}$. Furthermore, in Section 3.2.2 we suggested that the second derivative of phase $\tilde{\omega}'$ was a better indicator of phase nonlinearity. The variables ξ and χ represent two of the three components of $\tilde{\omega}'$ (3.10). The use of the final component τ is absent from the literature. If the second derivative of phase is valuable, its addition should improve classification. It may appear that the best feature simply would be the second derivative itself. This is not true. Consider the case where the first derivative is very large, but the second derivative is zero. Though phase may be linear, the constant of linearity is so large that we will incur errors due to phase wrapping. Consequently, at least one feature must restrict the size of ξ. The second derivative is not guaranteed to do this. The previously discussed features $|\xi|$ and $\sqrt{\xi^2 + \chi^2}$ both implement this restriction.

The next question is how to incorporate the second derivative as a feature. Again, the simple solution would be to use the features $|\xi|$ and $|\tilde{\omega}'|$. While this combination may work very well in theory, in practice it poses complications. Since ξ is a component of $\tilde{\omega}'$, the features are highly depen-

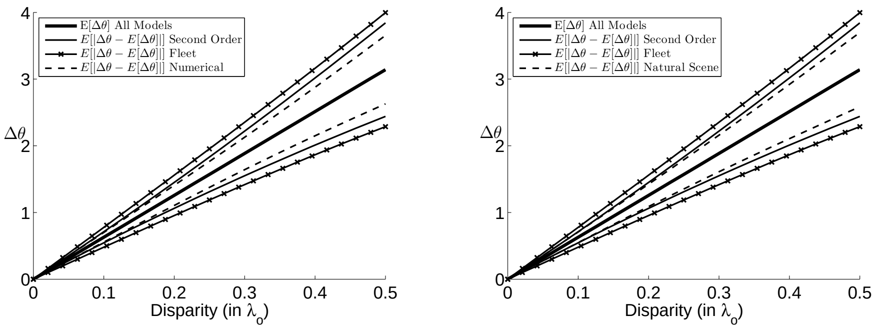


(a) (b)

Figure 3.2: Receiver operator characteristic (ROC) curves for Gabor kernels ($\omega_o = \pi/12$, $\beta = 0.8$) applied to images with constant disparity $\lambda_o/8$. The Y axis indicates the percentage of points with a disparity error of less than 25% that are correctly classified. The X axis indicates the percentage of points with a disparity error of less than 25% that are misclassified. 3.2(a) Results for white noise image. 3.2(b) Results for concatenated scan lines of natural scene.

dent. This results in an optimal decision region that is irregularly shaped, and consequently, not well approximated by applying a separate threshold to each feature. (As we said previously, this forms rectangular decision regions.) When thresholding it is important that the features be as independent as possible. For this reason we propose incorporating the second derivative by using the following features: $\sqrt{\xi^2 + \chi^2}$ and $|\tau|$.

Fig. 3.2 compares the use of three different feature sets. The first feature set uses the methodology most often advocated in the literature: $|\xi|$ and $|\chi|$. The second set uses $\sqrt{\xi^2 + \chi^2}$, altering the decision region from an rectangle to a circle. The final set uses both $\sqrt{\xi^2 + \chi^2}$ and $|\tau|$. Assuming that each feature



(a)

(b)

Figure 3.3: Comparison of mean phase difference $\mathrm{E}[\boldsymbol{\Delta\theta}]$ and expected deviation of phase difference about mean $\mathrm{E}[|\boldsymbol{\Delta\theta} - \mathrm{E}[\boldsymbol{\Delta\theta}]|]$ between simulated results, second order model, and model proposed by Fleet. The phase difference is measured at points satisfying $\sqrt{\xi^2 + \chi^2} < 1$. 3.3(a) Results for white noise image. 3.3(b) Results for concatenated scan lines of natural scene.

is thresholded at a certain value, Fig. 3.2(a) compares the resulting receiver operator characteristic curves for each set of features when applied to white noise with a constant disparity of $\lambda_o/8$. The class $\mathcal{C}_1$ represents the points we wish to detect and is defined as those points with a calculated disparity estimate within 25% of the true disparity. Class $\mathcal{C}_2$ contains all points not in $\mathcal{C}_1$. Approximately 96% of the points belong in class $\mathcal{C}_1$. Interestingly, the circular decision region $\sqrt{\xi^2 + \chi^2}$ slightly outperforms the rectangular region delineated by $|\xi|$ and $|\chi|$. This is remarkable because the space of possible decision regions for $|\xi|$ and $|\chi|$ has two degrees of freedom (length and width), while the circular decision region for $\sqrt{\xi^2 + \chi^2}$ has only one (radius). A fairer comparison might be to allow arbitrary elliptical regions.

Since the results in this dissertation were derived under the assumption of a white noise signal, the question of their extensibility remains. To address this concern we recreate Fig. 3.2(a), this time replacing the white noise image with concatenated scan lines of a natural scene. Obviously, these concatenated scan lines do not replicate the prospective transformations, occlusions, and other higher order effects that exist between two stereo images, but they do better represent the frequencies found in real images. Fig. 3.2(b) demonstrates that, though the overall performance of each set is slightly worse than with white noise, the order of efficacy remains the same. Additionally, the performance degradation for the feature set containing $\sqrt{\xi^2 + \chi^2}$ and $|\tau|$ is significantly less than that of the other feature sets.

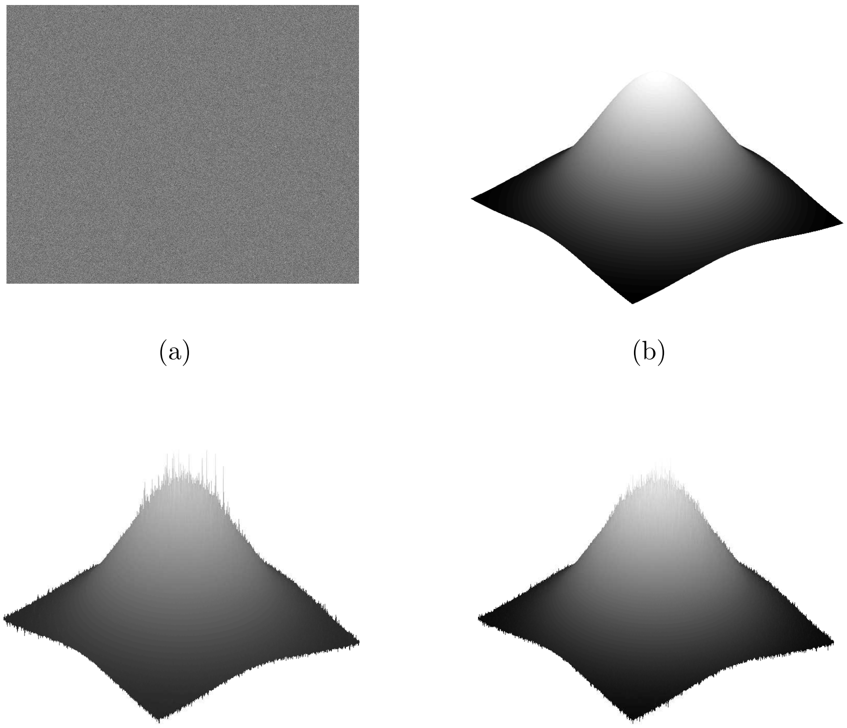
Another question about the extensibility of the results concerns the effects of the higher order terms in (3.2). Since we are making an assertion of linearity in a region by evaluating the second derivative at a point, we must question whether the higher order derivatives play an important role as we move farther away from this measurement. To address this concern we present Fig. 3.3 which illustrates various models of the mean phase difference $\mathrm{E}[\boldsymbol{\Delta\theta}]$ due to translation and the expected deviation about this mean $\mathrm{E}[|\boldsymbol{\Delta\theta} - \mathrm{E}[\boldsymbol{\Delta\theta}]|]$. Fig. 3.3 compares the simulated phase difference to both the second order model and the model presented by Fleet in [48]. The behavior of the second order model is given by

$$\mathrm{E}[\boldsymbol{\Delta\theta}\,(\Delta x)] = \mathrm{E}[\omega_o + \boldsymbol{\xi}]\Delta x + \frac{1}{2}\mathrm{E}[\boldsymbol{\tau} + 2\boldsymbol{\nu}]\Delta x^2 = \omega_o \Delta x \tag{3.50}$$

$$\begin{aligned}\mathrm{E}[|\boldsymbol{\Delta\theta}(\Delta x) - \mathrm{E}[\boldsymbol{\Delta\theta}(\Delta x)]|] =& \mathrm{E}\left[\left|(\omega_o + \boldsymbol{\xi})\Delta x + \frac{1}{2}(\boldsymbol{\tau} + 2\boldsymbol{\nu})\Delta x^2 - \omega_o \Delta x\right|\right] \\ \leq& \mathrm{E}[|\boldsymbol{\xi}|]\Delta x + \frac{1}{2}(\mathrm{E}[|\boldsymbol{\tau}|] + 2\mathrm{E}[|\boldsymbol{\nu}|])\Delta x^2. \qquad (3.51)\end{aligned}$$

The simulation and the models assume that measurements are made in regions of linear phase, i.e. at the measurement points we have have $\sqrt{\xi^2 + \chi^2} < \tilde{\rho}$, where $\tilde{\rho} = 1$. The threshold $\tilde{\rho} = 1$ is commonly advocated in the literature [61]. As we can see from the figure, both models adequately predict the simulated results for both the white noise and natural scene images.

With the goal of visually demonstrating the advantages of (3.39), we present a random dot stereo pair whose disparity is the isotropic Gaussian shown in Fig. 3.4(b). With a Gabor kernel ($\beta = 1$, $\omega_o = \pi/6$), disparity is estimated using Equation (3.3), producing the images in Fig. 3.4(c) and Fig. 3.4(d). Regions in Fig. 3.4(c) violating the criteria $\sqrt{\xi^2 + \chi^2} < 1.27$ were removed. Regions in Fig. 3.4(d) violating $\sqrt{\xi^2 + \chi^2} < 1.45$ or $|\tau| < 1.34$ were removed. Eliminated points were replaced by linear interpolation. To ensure a fair comparison, both sets of constraints were chosen so as to remove the same number of measurements (24%). The valid points remaining in both disparity maps are 95% identical. Still, the images shown in Figures 3.4(c) and 3.4(d) are visually different. This difference manifests as several isolated spikes in Fig. 3.4(c), which are both less frequent and less prominent in Fig. 3.4(d). The largest spikes occur at the peak of the Gaussian where the disparity is the greatest. This is understandable since the greater the disparity, the greater the error resulting from nonlinearities.



(a) (b)

(c) (d)

Figure 3.4: 3.4(a) Left random-dot stereo image. 3.4(b) Ideal disparity map. 3.4(c) Disparity map using $\sqrt{\xi^2 + \chi^2} < 1.27$. 3.4(d) Disparity map using $\sqrt{\xi^2 + \chi^2} < 1.45$ and $|\tau| < 1.34$.

Table 3.1 provides the mean squared error (MSE) between the true disparity map shown in Fig. 3.4(b) and the estimates shown in Figures 3.4(c) and 3.4(d). Since the majority of the disparity estimates are identical, it is instructive to calculate the MSE only considering the $N\%$ of the points with the largest squared error. These MSE calculations, evaluated for several values of N, are provided in the table. To futher confirm these results we replace

(a)

Top N%	Fig. 3.4(c) MSE	Fig. 3.4(d) MSE	Gain(%)
0.1	0.4560	0.3708	22.9896
1	0.2331	0.2101	10.9518
2	0.1861	0.1710	8.8463
5	0.1357	0.1276	6.3401
10	0.1053	0.1005	4.7955
15	0.0901	0.0865	4.1061
100	0.0383	0.0372	2.9586

(b)

Top N%	Fig. 3.4(c) SSIM	Fig. 3.4(d) SSIM	Gain(%)
0.1	0.9481	0.9725	2.5823
1	0.9799	0.9847	0.4869
2	0.9853	0.9881	0.2792
5	0.9907	0.9920	0.1281
10	0.9937	0.9944	0.0673
15	0.9951	0.9955	0.0441
100	0.9988	0.9988	0.0048

Table 3.1: Error metrics for disparity maps in Fig. 3.4. 3.0(a) MSE. 3.0(b) SSIM

MSE with the structural similarity index metric (SSIM) developed by Wang and Bovik [122, 124]. Shown to better model the human visual system than MSE, SSIM is a powerful metric for image quality assessment. Subsituting SSIM for MSE, Table 3.0(b) reprises Table 3.0(a). The SSIM values have a dynamic range of [0 : 1], with 1 signifying no error.

Both subtables in Table 3.1 indicate that as we consider fewer of the poorest estimates, the relative error between the disparity maps increases. This substantiates the conjecture made from visual inspection: the major difference between the disparity maps in Fig. 3.4(c) and Fig. 3.4(d) are the pronounced, but infrequent, spikes in Fig. 3.4(c). Since the SSIM metric at a point is a composite of the all the points in a region containing it, the derogating effects of singular spikes are mitigated, resulting in smaller relative error measures than the MSE metric. It is important to remember that SSIM was designed for the purpose of measuring image degradation in terms of visual appearance. From a purely visual perspective the occasional spikes do not significantly detract from our ability to extract visual information from the disparity map, perhaps explaining the similarity between the SSIM measures for both disparity maps.

3.5 Implementations

In this section we first review strategies for implementing a phase differencing registration algorithm and discuss its advantages and limitations. We then provide an overview of local correlation and its value in disparity estimation, comparing and contrasting its properties with those of phase differencing. We finally introduce a novel hybridization of the two that accentuates their positive aspects while mitigating their shortcomings.

3.5.1 Phase-Differencing

As we discussed in Section 3.2, phase differencing exploits the quasi-linear relationship between phase and translation to estimate disparity in a single step using the expression $\delta = \Delta\theta/\tilde{\omega}$, where $\tilde{\omega}$ is some approximation of $\frac{\partial}{\partial x}\theta$. This estimate can be refined in a Newton iteration

$$\delta^{(n+1)} = \frac{\theta_2\left(x + \delta^{(n)}\right) - \theta_1\left(x\right)}{\tilde{\omega}\left(x, \delta^{(n)}\right)}, \tag{3.52}$$

where $\tilde{\omega}(x, \delta)$ is judiciously chosen as the average instantaneous frequency over both images (though other formulations may suffice)

$$\tilde{\omega}(x, \delta) = \frac{1}{2}\left[\frac{\partial \theta_2}{\partial x}(x+\delta) + \frac{\partial \theta_1}{\partial x}(x)\right]. \tag{3.53}$$

Since phase wrapping forces $|\Delta\theta| < \pi$, in order to detect absolute disparities of size D we must have $\tilde{\omega} < \pi/D$. That is, high disparities must be detected at low frequencies. This suggests a coarse to fine process that allows large scales to bring the images into near registration before continuing to smaller scales.

In general, phase differencing schemes have both positive and negative properties. Perhaps the most positive property is the quasi-linearity of phase. This linearity allows us to rapidly estimate disparities to a very high resolution. Unfortunately, if at some level the requisite frequencies are not present in the image, the resulting specious measurements will cascade to lower levels. Additionally, image regions containing phase instabilities, identified using (3.13) and (3.39), preclude the use of (3.52). To produce a dense disparity map, estimates in these regions must be obtained using ancillary methods.

3.5.2 Local Correlation

Local correlation is another disparity estimation technique that relies heavily on phase information; though, unlike phase differencing, no explicit computation of phase is performed [47, 94]. To better elucidate this methodology consider the regular cross-correlation of the two images $i_1(x)$ and $i_2(x)$

and its representation in the frequency domain:

$$p(s) = \int i_1(x)\, i_2(x+s)\, dx = \frac{1}{2\pi} \int I_1(\omega)\, I_2^*(\omega)\, e^{-j\omega s} d\omega, \tag{3.54}$$

where $I_1(\omega)$ and $I_2(\omega)$ are the Fourier Transforms of $i_1(x)$ and $i_2(x)$, respectively. If $i_1(x)$ and $i_2(x)$ are related by a simple translation, $i_2(x) = i_2(x-\delta)$, then (3.54) reduces to

$$p(s) = \frac{1}{2\pi} \int |I_1(\omega)|^2 e^{-j\omega(s-\delta)} d\omega, \tag{3.55}$$

which obviously reaches its maximum when $s = \delta$. Removing the magnitudes from the previous expression (leaving only phase information) produces the phase-correlation

$$\tilde{p}(s) = \frac{1}{2\pi} \int e^{-j\omega(s-\delta)} d\omega = \delta(s-\delta), \tag{3.56}$$

where $\delta(\cdot)$ is the Dirac Delta function.

These results assume a global shift between images and must be spatially localized. Localization is achieved by first convolving each image with the i^{th} quadrature kernel $h_i(x)$

$$O_{1i}(x) = h_i(x) * i_1(x) = A_{1i}(x)\, e^{j\theta_{1i}(x)}, \tag{3.57}$$

$$O_{2i}(x) = h_i(x) * i_2(x) = A_{2i}(x)\, e^{j\theta_{2i}(x)}. \tag{3.58}$$

Usually, each $h_i(x)$ is tuned to a unique center frequency ω_i; and $O_{\cdot i}$ represents one channel of a bandpass decomposition of the image. Using (3.57) and (3.58), the local cross correlation can be expressed as

$$p(x,s) = \sum O_{1i}(x)\, O_{2i}^*(x+s) = \sum A_{1i}(x)\, A_{2i}(x+s)\, e^{j[\theta_{1i}(x) - \theta_{2i}(x+s)]}. \tag{3.59}$$

If the kernels are identically windowed complex exponentials with linearly spaced tuning frequencies, we can consider (3.57) and (3.58) as samples of the Short Time Fourier Transforms of $i_1(x)$ and $i_2(x)$. Recoving the disparity at x_o requires finding peaks in the real portion of $p(x_o, s)$ or zeros in the imaginary part. Since both are equivalent we can express the local correlation using only the real part as

$$p(x, s) = \sum A_{1i}(x) A_{2i}(x+s) \cos\left[\theta_{1i}(x) - \theta_{2i}(x+s)\right]. \tag{3.60}$$

In general, the values $A_{1i}(x)$ and $A_{2i}(x)$ can be considered weighting coefficients for each cosine function. Local correlation bases these weights on the magnitude of the filtered responses for each frequency channel. Phase-correlation assumes all weights are equivalent. In our experiments we do not normalize the coefficients, as we view the contributions of the magnitude information as beneficial. However, it is important to note that our approach (outlined in the following sections) does not depend on any specific choice of weights. This is not to suggest that the choice of weights is unimportant, but only that the following descriptions are independent of their selection.

Finally, we consider the good and bad aspects of local correlation, while comparing them with those of phase differencing. The most significant advantage of local correlation over phase differencing is its ability to cover the entire frequency spectrum. Secondly, phase wrapping is no longer a problem (or at least it is mitigated), allowing us to include high frequencies in the determination of large disparities. Unfortunately, these advantages come at the cost of dramatically increased computation time. Furthermore, a direct implementation would seem to require sampling $p\left(x,s\right)$, tying the resolution to the sampling density. Phase-differencing has no theoretical upper bound on its resolution.

3.5.3 Hybrid Approach

In this section we introduce a novel combination of phase differencing and local correlation that retains the advantages of each, but without the normally attendant complications. We begin by incorporating the most important aspect of phase differencing, linear phase, into the local correlation model as follows:

$$p\left(x,s\right) = \sum A_{1i}\left(x\right) A_{2i}\left(x\right) \cos\left[\tilde{\omega}_i\left(x\right)\left(s - \delta_i\left(x\right)\right)\right], \tag{3.61}$$

where $\delta_i\left(x\right) = \frac{1}{\tilde{\omega}_i}\left[\theta_{2i}\left(x\right) - \theta_{1i}\left(x\right)\right]$ is the disparity found by phase differencing. For a given point x_o, the function $p\left(x_o,s\right)$ becomes a simple sum of cosine waves. Equation (3.61) can be seen as a weighted sinusoidal voting for the best disparity, where the weights are the products of the magnitudes. This sinusoidal voting establishes a framework for reconciling disparities across multiple frequencies. If there is only one frequency channel then the maximums in s exactly correspond to the choice of disparities that would result from phase differencing (in phase differencing the maximum closest to zero is selected). As additional frequencies are added, the occurrence of global maximums becomes less and less frequent. Within the acceptable range of disparities, we hope to ultimately reduce this number to one. Furthermore, for a given $p\left(x_o,s_o\right)$, a confidence value (similar to that proposed in [47]) can be realized by normalizing the correlation by the sum of the weights over all frequency channels

$$c\left(x,s\right)=\frac{\sum A_{1i}\left(x\right)A_{2i}\left(x\right)\cos\left[\tilde{\omega}_i\left(x\right)\left(s-\delta_i\right)\right]}{\sum A_{1i}\left(x\right)A_{2i}\left(x\right)}. \tag{3.62}$$

When all the disparities δ_i coincide, the confidence reaches its maximum value of 1. As they begin to separate, the confidence degrades. Finally, it is important to note that (3.61) is only valid when phase is relatively linear with respect to translation, i.e. regions satisfying (3.13) and (3.39). For frequencies channels outside regions of linearity, we would choose to exclude their respective cosine functions from (3.61) and (3.62).

For an image point x_o, the goal is to find the value of s (within a predetermined range of disparities) that maximizes (3.61). Since $p\left(x_o,s\right)$ will generally have many local maximums, initially we will need to perform a global search. Once we are within the domain of convergence of the global maximum, we can proceed with a Newton iteration. First consider the iterative scheme implied by the following equations:

$$s^{(n+1)}=s^{(n)}+s_r^{(n)} \tag{3.63}$$

where, $s_r^{(n)}$ represents the residual disparity that maximizes

$$p\left(x_o,s^{(n)}\right)=\sum A_{1i}\left(x_o\right)A_{2i}\left(x_o+s^{(n)}\right)\cos\left(\tilde{\omega}_i^{(n)}\left[s_r^{(n)}-\delta_i^{(n)}\right]\right), \tag{3.64}$$

and $\delta_i^{(n)}$ and $\tilde{\omega}_i^{(n)}$ represent each channel's disparity estimate and instantaneous frequency for the current iteration. They are found as follows:

$$\delta_i^{(n)}=\frac{\theta_{2i}\left(x_o+s^{(n)}\right)-\theta_{1i}\left(x_o\right)}{\tilde{\omega}_i^{(n)}} \tag{3.65}$$

$$\tilde{\omega}_i^{(n)} = \frac{1}{2}\left[\frac{\partial}{\partial x}\theta_{2i}\left(x_o + s^{(n)}\right) - \frac{\partial}{\partial x}\theta_{1i}\left(x_o\right)\right] \tag{3.66}$$

The global search coarsely samples (3.64) and then locates the maximum $s_r^{(n)}$. Once $s_r^{(n)} < \pi/\omega_N$, where ω_N is the highest center frequency, we know the algorithm has converged into a region in which even the highest frequency has only one maximum. We can now begin the high fidelity Newton iteration. (As a side note, if during the global search the instantaneous frequency $\tilde{\omega}_i^{(n)}$ in (3.64) and (3.65) is approximated by the center frequency ω_i of the quadrature filter, the cosines can be precomputed, greatly reducing the computational complexity. Our computer implementation uses this approximation.)

The Newton iteration requires the calculation of the derivative of (3.64). In order to derive an expression for this derivative we first rewrite (3.64) in the following more manageable form:

$$f\left(s\right) = \sum a_i \cos\left(\tilde{\omega}_i\left[s - \delta_i\right]\right). \tag{3.67}$$

We now determine its first derivative and then approximate it using a second order Taylor series in s

$$\begin{aligned} f'\left(s\right) &= \sum -a_i\tilde{\omega}_i \sin\left(\tilde{\omega}_i\left[s - \delta_i\right]\right) \\ &\approx \sum \left[a_i\tilde{\omega}_i \sin\left(\tilde{\omega}_i\delta_i\right) + a_i\tilde{\omega}_i^2 \cos\left(\tilde{\omega}_i\delta_i\right) s\right]. \end{aligned} \tag{3.68}$$

Without this Taylor expansion of $f'\left(s\right)$, the equation $f'\left(s\right) = 0$ is transcendental and must be solved iteratively. Using the second order approximation of $f'\left(s\right)$, we can set $f'\left(s\right) = 0$ and easily solve for s. This solution followed by further approximating the trigonometric functions by a second order Taylor series in δ_i we have

$$s = \frac{\sum a_i \tilde{\omega}_i \sin\left(\tilde{\omega}_i \delta_i\right)}{\sum a_i \tilde{\omega}_i^2 \cos\left(\tilde{\omega}_i \delta_i\right)} \approx \frac{\sum a_i \tilde{\omega}_i^2 \delta_i}{\sum a_i \tilde{\omega}_i^2}. \tag{3.69}$$

We now have a new means for finding $s_r^{(n)}$

$$s_r^{(n)} = \frac{\sum A_{1i}\left(x_o\right) A_{2i}\left(x_o + s^{(n)}\right) \left[\tilde{\omega}_i^{(n)}\right]^2 \delta_i^{(n)}}{\sum A_{1i}\left(x_o\right) A_{2i}\left(x_o + s^{(n)}\right) \left[\tilde{\omega}_i^{(n)}\right]^2}. \tag{3.70}$$

Interestingly, (3.70) estimates the overall disparity by performing a weighted average of the disparities at each frequency channel. Since the individual disparity estimates are linear with respect to phase, so is their summation in (3.70). Consequently, we can expect a convergence rate similar to that of typical phase differencing.

Just as with phase differencing, the hybrid approach uses a coarse to fine process. The coarseness is a function of the spatial extent of filters $h_i\left(x\right)$. Since filters with larger spatial extent are able to detect higher disparities, they first help bring the images into near registration before applying the next smaller filters. Since spurious estimations (caused by occlusions etc.) could cascade down the levels of the pyramid, regularization should be performed following each scale. This is the topic of the next section.

3.5.4 Regularization

Due to situations such as foreshortening and occlusions, registration is an ill-posed problem with solutions that are nonexistent or not unique [97].

The introduction of a smoothness constraint can make the problem well posed. The process of incorporating this constraint is called regularization [5, 51, 97]. Though previous derivations were made clearer using one-dimensional signals, this section is best explained with two-dimensional images.

Before discussing the regularization process, we first address the need to convert the confidence measures in (3.62) into a more suitable form. Currently, each confidence value is assessed independently of every other. We propose that these measures would have more meaning if they were viewed relative to the general distribution of all confidence measures. Consider the case where 50% of confidence measures are c_1 an 50% are c_2. Consider another case where 99% of confidence measures are c_1 an 1% are c_2. In the first case, the confidence value c_1 is part of the normal process inherent in the disparity map. In the second case, c_1 appears more likely to be outlier. We propose that the weights signifying our "certitude" about the disparity estimates must reflect the general appearance of the confidence measures throughout the entire image.

If $\boldsymbol{c}$ is considered a random variable, then the distribution of $1 - \boldsymbol{c}$ can be well modeled as an exponential. We can robustly measure this distribution's mean μ (the one sufficient statistic of the exponential distribution) by recognizing

$$\mu = \frac{\text{median}\,[\boldsymbol{c}]}{\ln(0.5)}.$$

Using this statistic with the cumulative distribution function of the exponential, we arrive at the expression converting confidence c into relative confidence

$\tilde{c}$

$$\tilde{c}\left(c\right) = e^{-\frac{1-c}{\alpha\mu}},$$

where $\alpha > 0$ represents the degree of certainty we require. The larger the value of α, the more outliers are tolerated. In words, this equation reflects the probability that a confidence value of c or larger would be produced by an exponential distribution whose mean is α times the statistically estimated mean μ.

Thresholding the relative confidence measures, we are able identify disparity estimates that should be removed. Each of these estimates is replaced with a weighted sum of its disparity estimate and those of its neighbors. This weighting is best expressed as a smoothing enacted on the entire image

$$\delta_\sigma\left(x\right) = \frac{\delta\left(x,y\right)\tilde{c}\left(x,y\right) * g_\sigma\left(x,y\right)}{\tilde{c}\left(x,y\right) * g_\sigma\left(x,y\right)}, \tag{3.71}$$

where $g_\sigma\left(x\right)$ is a Gaussian filter with standard deviation σ and $\tilde{c}\left(x,y\right)$ are the relative confidence values. Once we have removed the most spurious points, we can perform standard regularization. The goal of regularization is to find the function $\delta_r\left(x,y\right)$ that minimizes [57]

$$\int \tilde{w}\left(x,y\right)\left(\delta_r\left(x,y\right) - \delta\left(x,y\right)\right)^2 + \lambda\left(\frac{\partial^2}{\partial x^2}\delta_r\left(x,y\right) + \frac{\partial^2}{\partial y^2}\delta_r\left(x,y\right)\right)^2 dxdy. \tag{3.72}$$

The first term in (3.72) represents the error between the computed disparity map $\delta\left(x,y\right)$ and the regularized map $\delta_r\left(x,y\right)$. The second term measures the smoothness of $\delta_r\left(x,y\right)$ in terms of the sum of its second derivatives. The second derivative seems more appropriate than the often used gradient, since a constant slope is not undesirable. The parameter λ determines the degree of smoothing. Solving this functional requires the calculus of variations and is beyond the scope of this dissertation. Instead, we will turn to the discrete form of (3.72):

$$\sum \left[w_{ij} \left(u_{ij} - \delta_{ij} \right)^2 + \lambda \left(u_{ij} - \bar{u}_{ij} \right)^2 \right] dx, \tag{3.73}$$

where $\bar{u}_{ij}$ is the four-neighbor average of the points surrounding u_{ij}. (Extra constants were absorbed into λ.) In an effort to maximize (3.73), we differentiate it with respect to u_{ij} and set the resultant expression equal to zero. For an $M \times N$ image, the exact solution requires solving MN linear equations. Instead, we employ a Gauss-Seidel relaxation process

$$u_{ij}^{(n+1)} = \frac{w_{ij}\delta_{ij} + \lambda \bar{u}_{ij}^{(n)}}{w_{ij} + \lambda}. \tag{3.74}$$

Once the iterations converge, the smoothed disparity map can be used as the starting point for the next finer scale.

3.5.5 Experiments

We implemented the previously discussed hybrid technique using $N = 20$ Gabor kernels

$$h_i(x) = \frac{1}{\left(\sqrt{\pi}\sigma_g\right)^{\frac{1}{2}}} e^{-\frac{x^2}{2\sigma_g^2}} e^{j\omega_i x}, \tag{3.75}$$

at each scale in the coarse to fine process. At the coarsest scale for the pentagon images (see Fig. 3.5) the center frequencies ω_i of each of the N kernels were spaced linearly, covering the range $[\pi/16,\ 15\pi/16]$. All filters had identical bandwidths of $\sigma_\omega = \pi/48$. Three levels were used, with each step increasing the resolution by a factor of two, i.e. for the second level the center frequencies covered $[\pi/8,\ 7\pi/8]$ and $\sigma_\omega = \pi/24$. The stereo images and resultant disparity map are shown in Fig. 3.5. Other than the addition of one coarser scale, the baseball, White House, and Venus images were processed identically. Figure 3.7, Fig. 3.9, and Fig. 3.11 illustrate their stereo images and corresponding disparity maps.

Additionally, it is important to compare the conclusions drawn in this paper to other prevalent methods. The first algorithm for comparison is the coarse-to-fine approach advocated by Fleet [49]. Fleet uses large wavelengths to bring the images into approximate registration before continuing to finer scales. Fleet uses a single frequency at each each level of the coarse-to-fine pyramid. This differs from our method which allows multiple frequencies to be used at each level. The second algorithm used for comparison was developed by Ouali [87]. Similar to our algorithm, Ouali is able to extract disparity from multiple frequencies. The difference lies in the manner of combination. Ouali defines the “best” disparity measurement as that which corresponds to the frequency with the greatest amplitude response. This differs significantly from our sinusoidal combination of disparity measurements.

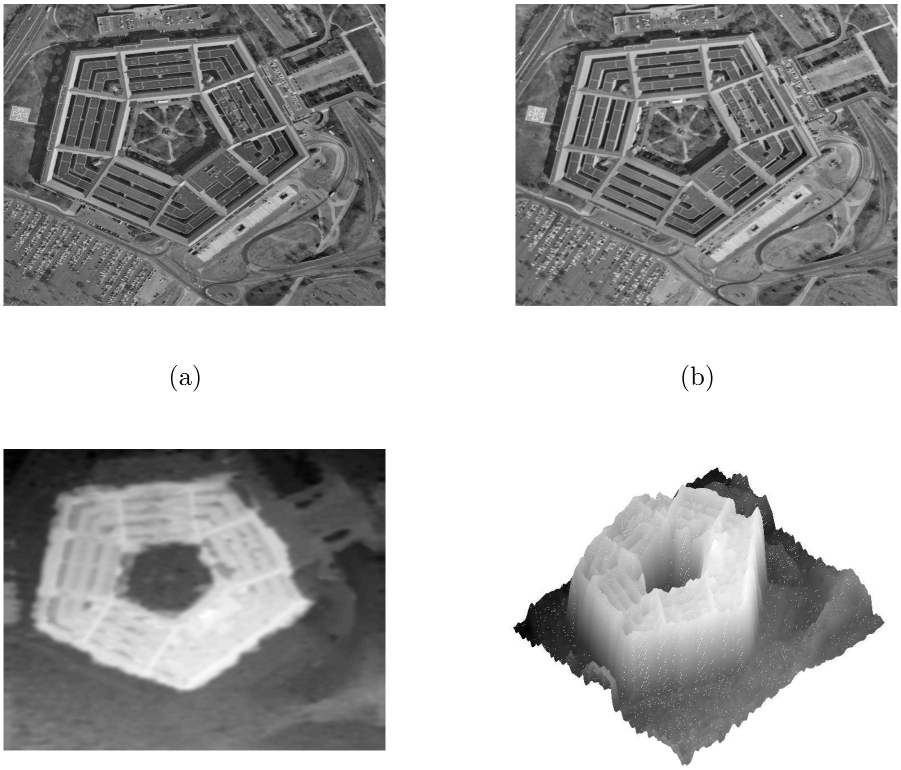
The novel elements introduced in this paper are twofold: the improved detection of nonlinearities and the sinusoidal combination of disparity estimates over multiple frequencies. Since comprehensive stereo algorithms are amalgamations of several complex components, to conduct a fair comparison between our algorithm and those of Fleet and Ouali, the correspondence algorithms will differ only with respect to these two elements. First our nonlinearity detection is replaced by the methods prevalent in the literature. Second our sinusoidal method for synthesizing disparity estimates is replaced with the methods proposed by Fleet and Ouali. For the Fleet algorithm a single frequency is used to procure a single disparity estimate. For the Ouali algorithm the disparity measurements attained over multiple frequencies are combined using his selection criteria. Other than these differences, the algorithms are completely identical.

The comparison results are shown in Fig. 3.6, Fig. 3.8, Fig. 3.10, and Fig. 3.12. The differences are visually striking. It is obvious that the Fleet algorithm tends to miss fine detail. This is a result of using only a single frequency to obtain a disparity estimate. Additionally, the Fleet algorithm has places of large errors possibly resulting from the lack of that specific frequency. The Ouali algorithm tends to retain more detail than the Fleet algorithm, but it also is far more susceptible to noise. Our algorithm appears to qualitatively outperform both for all the example images.

3.6 Conclusion

In this chapter we were able to derive probability density functions for the components of both the first and second derivatives of localized phase. We quantified the effects on these pdfs caused by constraining their domains to regions of linear phase. In order to identify these regions, we proposed using the second derivative of phase. We then described how prevalent methods proposed in the literature can be seen as approximations of the second derivative. We empirically validated these assertions and their extensibility by comparing the different methods on white noise, a natural scene, and a random-dot stereo pair.

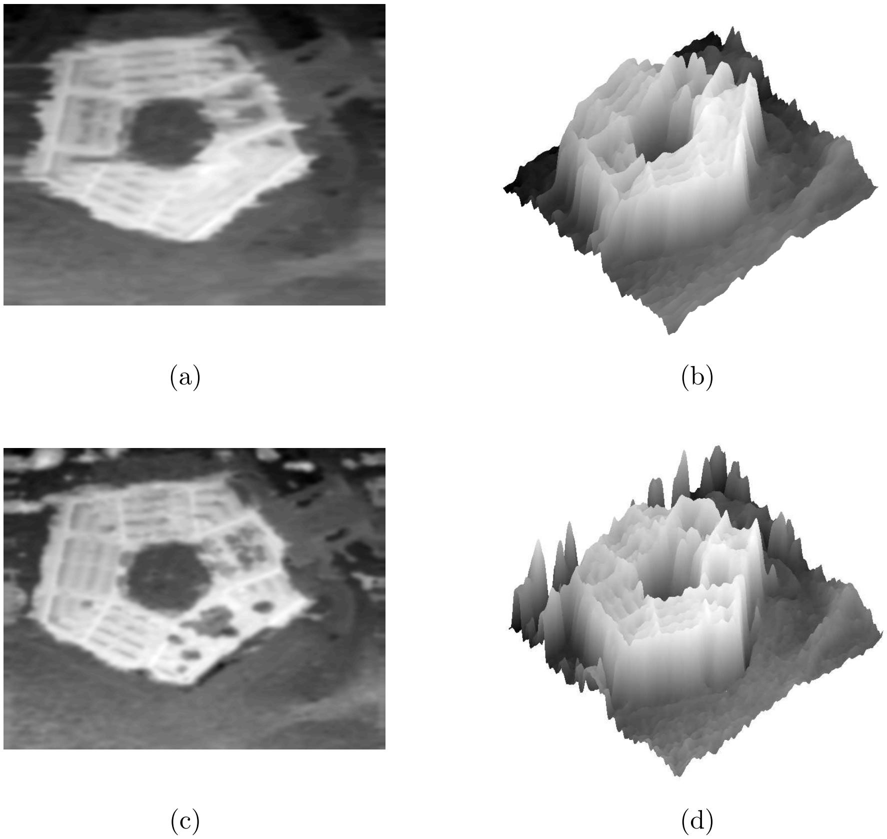
We also addressed the difficulty of multiscale fusion in phase differencing by introducing a novel registration technique that incorporates strategies of both phase differencing and local correlation. This hybrid approach allowed us to combine the rapidly calculated, high resolution results associated with phase differencing with the multiscale information available to local correlation. Finally, we demonstrated the performance on four pairs of stereo images, comparing its behavior to those of two other prominent algorithms.



(a) (b)

(c) (d)

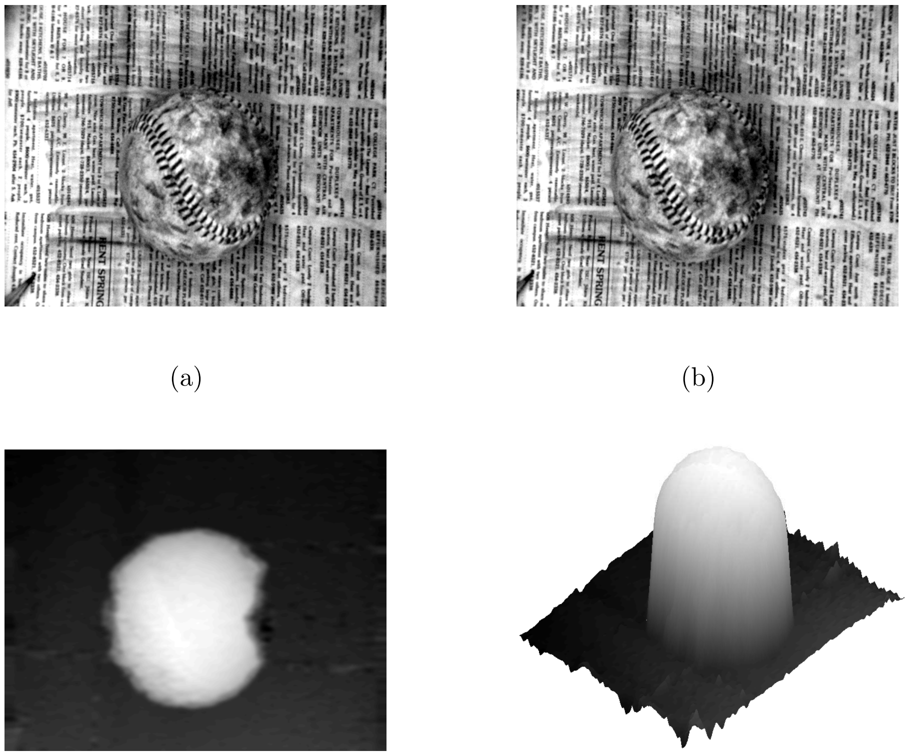
Figure 3.5: Pentagon images with resultant disparity maps from our algorithm. 3.5(a) Left image of pentagon. 3.5(b) Right image of pentagon. 3.5(c) Disparity map. 3.5(d) Surface plot of disparity map.



(a) (b)

(c) (d)

Figure 3.6: Pentagon disparity maps for algorithms by Fleet and Ouali. 3.6(a) Fleet disparity map. 3.6(b) Fleet surface plot of disparity map. 3.6(c) Ouali disparity map. 3.6(d) Ouali surface plot of disparity map.



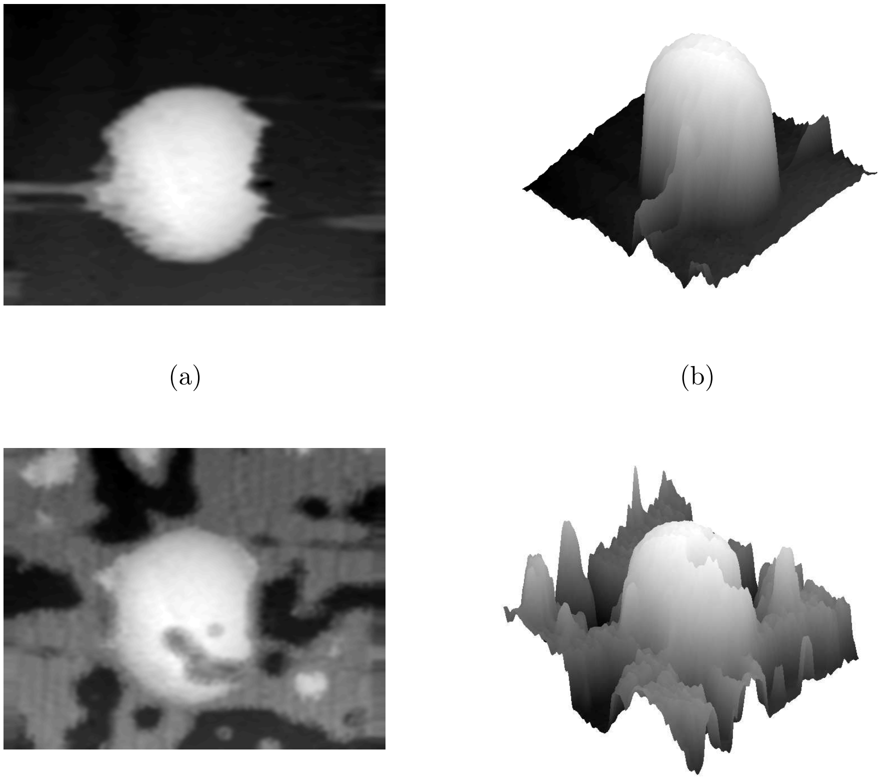
(a)

(b)

(c)

(d)

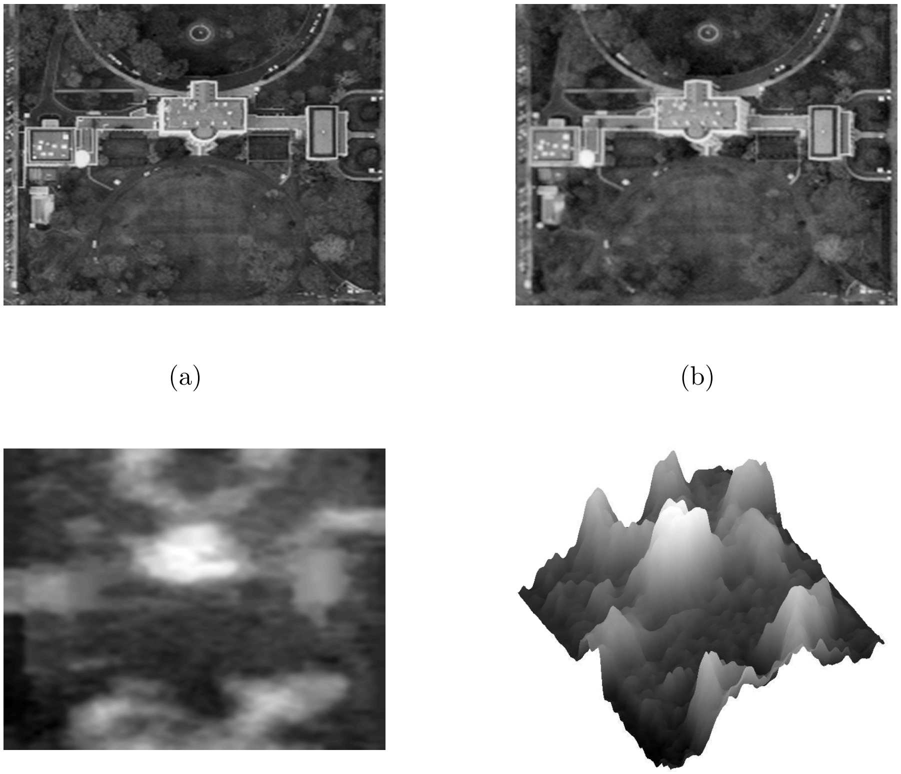
Figure 3.7: Baseball images with resultant disparity maps from our algorithm. 3.7(a) Left image of baseball. 3.7(b) Right image of baseball. 3.7(c) Disparity map. 3.7(d) Surface plot of disparity map.



(a) (b)

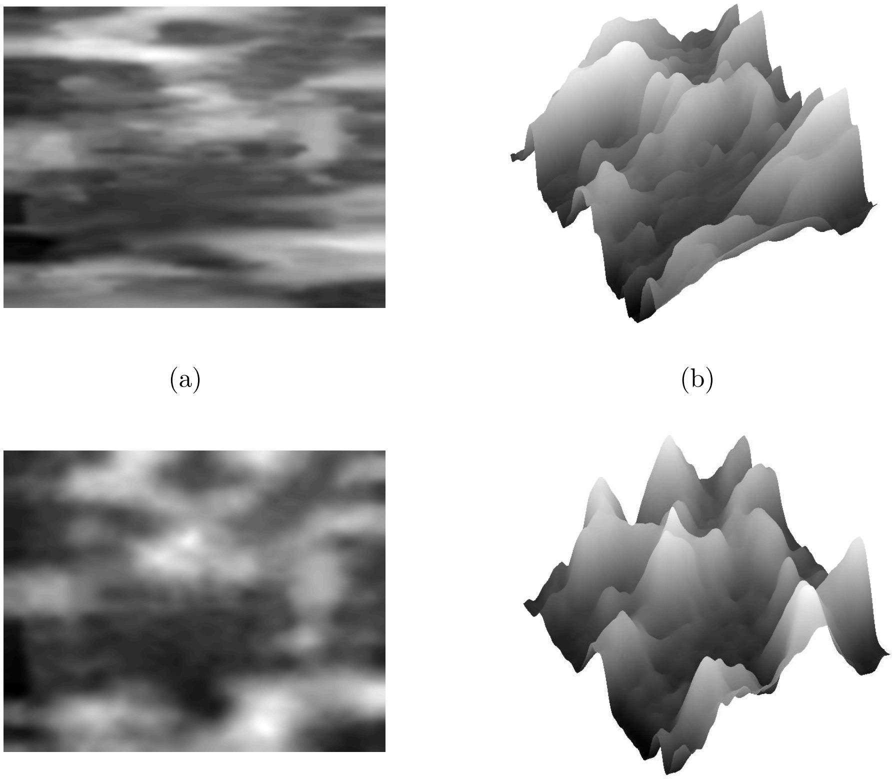
(c) (d)

Figure 3.8: Baseball disparity maps for algorithms by Fleet and Ouali. 3.8(a) Fleet disparity map. 3.8(b) Fleet surface plot of disparity map. 3.8(c) Ouali disparity map. 3.8(d) Ouali surface plot of disparity map.



(a) (b) (c) (d)

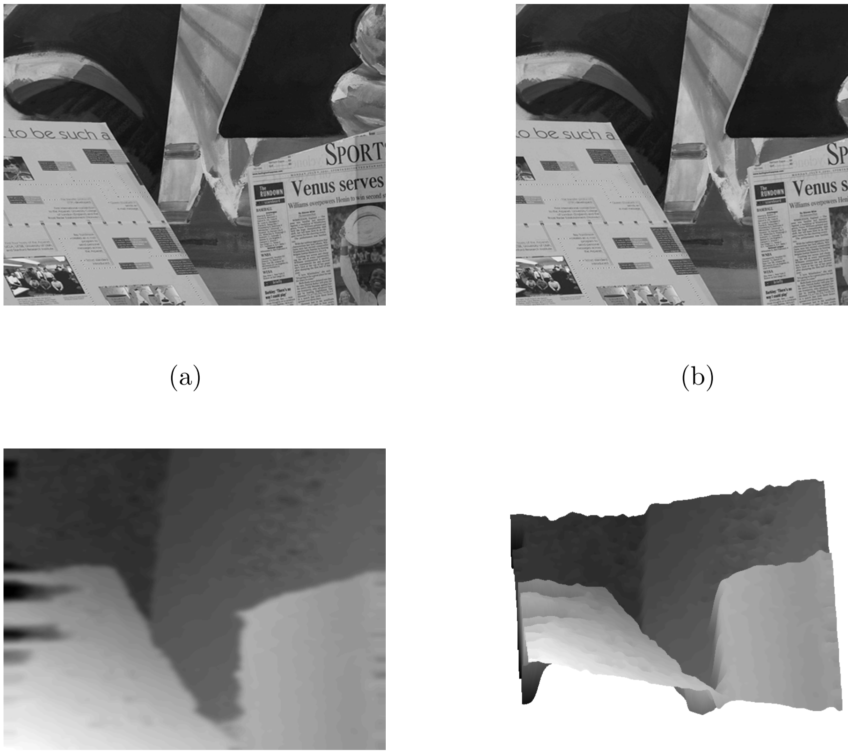
Figure 3.9: White House images with resultant disparity maps from our algorithm. 3.9(a) Left image of White House. 3.9(b) Right image of White House. 3.9(c) Disparity map. 3.9(d) Surface plot of disparity map.



(a) (b)

(c) (d)

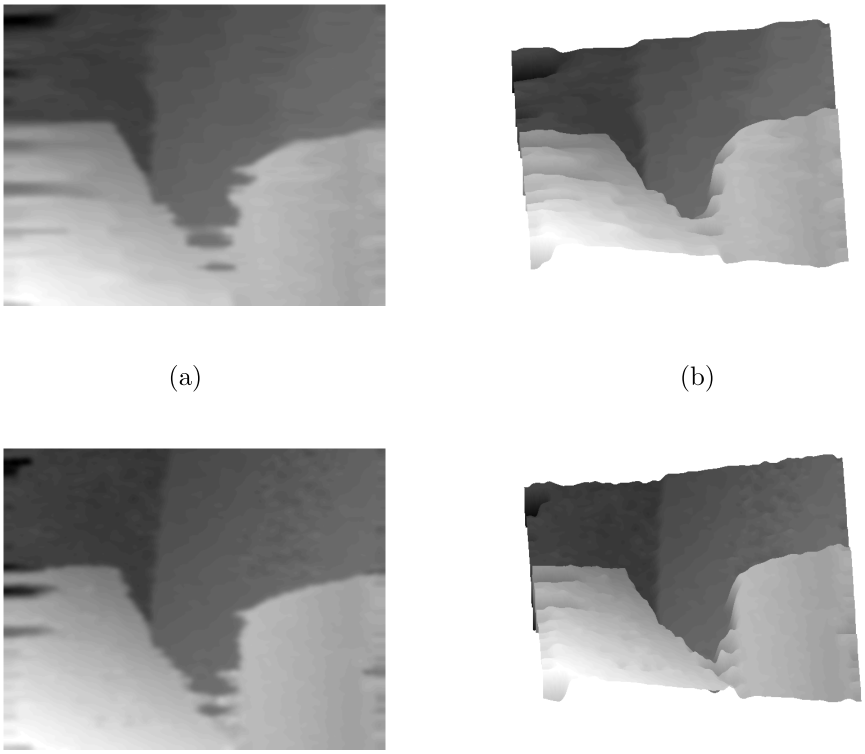
Figure 3.10: White House disparity maps for algorithms by Fleet and Ouali. 3.10(a) Fleet disparity map. 3.10(b) Fleet surface plot of disparity map. 3.10(c) Ouali disparity map. 3.10(d) Ouali surface plot of disparity map.



(a) (b)

(c) (d)

Figure 3.11: Venus images images with resultant disparity maps from our algorithm. 3.11(a) Left image of White House. 3.11(b) Right image of White House. 3.11(c) Disparity map. 3.11(d) Surface plot of disparity map.



(a)

(b)

(c)

(d)

Figure 3.12: Venus disparity maps for algorithms by Fleet and Ouali. 3.12(a) Fleet disparity map. 3.12(b) Fleet surface plot of disparity map. 3.12(c) Ouali disparity map. 3.12(d) Ouali surface plot of disparity map.

Chapter 4

Epipolar Spaces and Optimal Sampling Strategies

4.1 Introduction

4.1.1 Sampling Strategies and Registration

In the previous chapter we discussed the first of two primary components of depth recovery for binocular active vision systems: registration. We demonstrated how registration algorithms can use local phase to successfully correspond images. In this section we consider the second component: sampling. How an image is sampled can directly effect both the speed and accuracy of the registration process. Additionally, certain sampling schemes require specific registration algorithms. Since the results of this section will encourage the use of nonstandard sampling strategies, it is important that we consider the availability of registration techniques suitable for arbitrarily sampled images. Fortunately, phase based registration algorithms are extremely well suited for registration of such images. The adaption of phase based algorithms for stereo correspondence of nonuniformly sampled images is a topic discussed in the next chapter.

4.1.2 Sampling Strategies and Active Binocular Vision Systems

Active binocular vision systems (ABVS) provide a powerful means for extracting information from a complex scene. With a virtually unlimited field of view they have access to huge amounts of information, yet are able to confine their resources to specific regions of interest. Additionally, their ability to actively analyze their environment enables them to address, in a well-posed manner, tasks that may be ill-posed for a passive observer [4].

ABVS are especially well suited for the recovery of depth information. Depth recovery requires registering the two views, a notoriously difficult problem. This task can be greatly simplified if the camera geometry is known. With geometric knowledge of the stereo configuration, the search for corresponding points can be restricted to epipolar lines. Ascertaining the actively changing stereo geometry requires calibration. This calibration procedure can be highly complex, involving motorized lens calibration, kinematic calibration, and head/eye calibration [111]. Consequently, mechanical calibration requires highly accurate, often expensive equipment. For some ABVS, such as those deployed for planetary exploration using mobile rovers, calibration may not be feasible [135].

While it is possible to register images without calibration information, such unconstrained algorithms are usually far more time consuming and prone to error. In this chapter we propose a compromise. Even without knowledge of the exact stereo geometry we can restrict the region of correspondence by imposing limits on the possible range of configurations. That is, by restricting the range of vergence angles, baseline distances, and focal lengths etc., we can confine our search for matching points to what we refer to as *epipolar spaces*. For each point in one image, we define the corresponding epipolar space in the other image as the union of all associated epipolar lines over all possible system geometries.

Epipolar spaces eliminate the need for calibration at the cost of an increased search region. One approach to mitigate this increase is the application of a space variant sampling or foveation strategy. The application of such a strategy to stereo vision tasks is not new. Space variant transformations such as log-polar [127], reciprocal wedge transform [71], foveated wavelet transform [125], and fish-eye [110] have been successfully applied to binocular vision problems such as vergence [74], time-to-impact analysis [114], and depth recovery of a scene [23, 103]. Yet, in these instances, the sampling strategies were introduced elsewhere and later applied to areas of stereo vision. Only rarely has a foveation strategy been specifically tailored for a stereo vision task. Both Basu [9] and Elnagar [42] derived optimal sampling schemes with respect to the error discretization of depth measurements. Klarquist and Bovik [70] designed a real-time foveated stereo technique that adapts to the specific geometry, always producing horizontal epipolar lines.

In this dissertation we first formalize the concept of epipolar spaces. We then discuss the general mathematics of applying nonuniform sampling strategies to epipolar spaces. Finally, we present an optimal sampling scheme specifically designed to minimize the average area per epipolar space.

4.2 Epipolar Spaces

Knowledge of the camera geometry can be extremely valuable in the registration process, reducing the search for matching points to epipolar lines. Consider the stereo geometry in Fig. 4.1. Here two pinhole cameras whose optical centers are located at C_l and C_r converge at the fixation point V. Both cameras have identical focal lengths f. The left and right camera rotation angles are θ_l and θ_r. Each camera has an associated right-hand coordinate system originating at its optical center. If $\tilde{m}_l = [x_l, y_l, z_l, 1]^t$ and $\tilde{m}_r = [x_r, y_r, z_r, 1]^t$ are projective world coordinates in the left and right camera frames, respectively, then their relation is $\tilde{m}_r = [R\,|\,T]\,\tilde{m}_l$, where

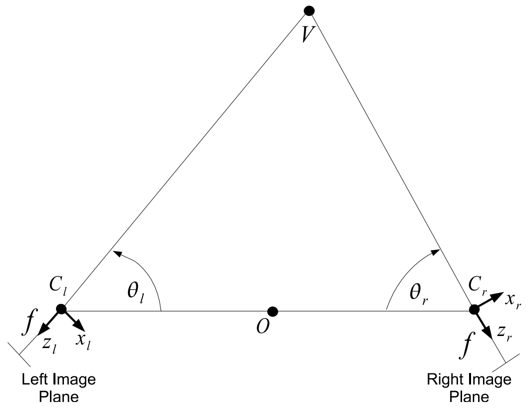


Figure 4.1: Stereo geometry.

$$R = \begin{bmatrix} -\cos(\theta_l+\theta_r) & 0 & -\sin(\theta_l+\theta_r) \\ 0 & 1 & 0 \\ \sin(\theta_l+\theta_r) & 0 & -\cos(\theta_l+\theta_r) \end{bmatrix} \tag{4.1}$$

is the rotation matrix and

$$T = \begin{bmatrix} -\sin(\theta_r) \\ 0 \\ -\cos(\theta_r) \end{bmatrix} \tag{4.2}$$

is the translation vector. Let $m_l = [u_l, v_l, 1]^t$ and $m_r = [u_r, v_r, 1]^t$ be projective coordinates in the left and right image planes. It is well known that the equation relating corresponding epipolar lines is

$$m_r^t A_r^{-t} T_\times R A_l^{-1} m_l = 0, \tag{4.3}$$

where, in our simplified geometry,

$$A_l = A_r = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.4}$$

are the identical intrinsic matrices and

$$T_\times = \begin{bmatrix} 0 & \cos(\theta_r) & 0 \\ -\cos(\theta_r) & 1 & \sin(\theta_r) \\ 0 & -\sin(\theta_r) & 0 \end{bmatrix} \tag{4.5}$$

implements the crossproduct as a matrix. The matrix $E = T_\times R$ is the essential matrix and relates the coordinate frames. The matrix $F = A_r^{-t} E A_l^{-1}$ is the fundamental matrix [46] and includes the intrinsic parameters of the cameras. The expression in (4.3) can be simplified to the following:

$$v_l = v_r \frac{f \sin(\theta_l) + u_l \cos(\theta_l)}{f \sin(\theta_r) - u_r \cos(\theta_r)}. \tag{4.6}$$

That is, for a given point (u_r, v_r) in the right image, the matching point in the left image (if not obscured) lies on the line given by (4.6).

We now consider the situation where the camera configuration actively changes and we no longer know the specific geometry. In such a situation we will not be able to restrict our search for corresponding points to epipolar lines. However, even though we may not know the precise values of parameters such as focal length, baseline distance, and camera rotation angles, we can establish acceptable ranges for these values. Consequently, we can still restrict the location of corresponding points across images. For a given point in one image, the matching point in the other is confined to a region defined by the union of all corresponding epipolar lines produced over all possible camera configurations. We call these continuous regions *epipolar spaces.*

The goal of the remainder of this section is to quantify these epipolar spaces for a stereo rig with a fixed baseline and fixed focal length as shown in Fig. 4.1. In this configuration the only variable parameters that effect the epipolar geometry are the rotation angles θ_l and θ_r. Translation and rotation of the entire stereo rig about O, while allowed, do not influence the epipolar geometry. We establish the range of rotation angles by confining them to the interval

$$\theta_l, \theta_r \in \left[\theta_M, \pi - \theta_M\right], \tag{4.7}$$

where θ_M is the minimum angle relative to the baseline.

Although theoretically a matching point can lie anywhere on the corresponding epipolar lines, the search is usually restricted to a maximum horizontal disparity. Bounding the horizontal disparity has the effect of limiting the depth around the horopter at which objects can be fused. For our purposes, we assume a maximum horizontal disparity defined by

$$|d| = |u_l - u_r| \leq D. \tag{4.8}$$

For the right image point I_r Fig. 4.2 illustrates several corresponding epipolar lines in the left image. Each separate epipolar line results from a unique geometric configuration, i.e. a unique combination of θ_l and θ_r The dashed vertical lines delimit the maximum allowable horizontal disparities defined by (4.8).

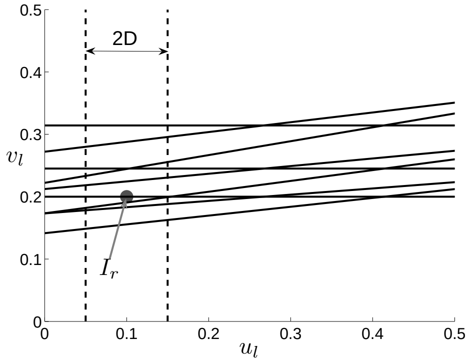


Figure 4.2: Left image epipolar lines corresponding to right image point I_r for several unique geometric configurations. Specifically, the camera rotation angles θ_l and θ_r are each set to an element of $\{\pi/4, \pi/3, \pi/2\}$, producing nine combinations.

The restriction imposed by (4.8) determines the leftmost and rightmost bounds of the epipolar spaces. The upper and lower bounds are determined by maximizing and minimizing (4.6) with respect to both θ_l and θ_r, respectively. The maximum can be found by separately maximizing the numerator and minimizing the denominator. Taking the derivative of the numerator with respect to θ_l, setting it to zero, and solving for θ_l yields the maximizing value $\theta_l = \operatorname{atan}(f/u_l)$. The denominator (which must be positive so long as the cameras never image each other, i.e. the images do not contain the epipoles) is minimized when $\theta_r = \theta_M$. Inserting these results into (4.6), the upper bound becomes

$$v_{l,max} = v_r \frac{\sqrt{f^2 + u_l^2}}{f \sin(\theta_M) - u_r \cos(\theta_M)}. \tag{4.9}$$

The minimization of (4.6) is performed similarly with $\theta_l = \pi - \theta_M$ and $\theta_r = \operatorname{atan}(-f/u_l)$, producing

$$v_{l,min} = v_r \frac{f \sin(\theta_M) - u_r \cos(\theta_M)}{\sqrt{f^2 + u_l^2}}. \tag{4.10}$$

Since the horizontal extents of these bounds are limited by (4.8), $f^2+u_l^2$ in the preceding equations can be rewritten as $f^2 + (u_r+\epsilon)^2$, where $\epsilon \in [-D, D]$. In practice $f \gg \epsilon$ and $f \geq u_r$, allowing the following approximation: $f^2+u_l^2 \approx f^2+u_r^2$. Incorporating this result into (4.9) and (4.10) produces

$$v_{l,max} \approx v_r \frac{\sqrt{f^2 + u_r^2}}{f \sin(\theta_M) - u_r \cos(\theta_M)} = v_r c(u_r) \tag{4.11}$$

$$v_{l,min} \approx v_r \frac{f \sin(\theta_M) - u_r \cos(\theta_M)}{\sqrt{f^2 + u_r^2}} = \frac{v_r}{c(u_r)}, \tag{4.12}$$

where

$$c(u_r) = \frac{\sqrt{f^2 + u_r^2}}{f \sin(\theta_M) - u_r \cos(\theta_M)}. \tag{4.13}$$

Remarkably, an epipolar space is well modeled by a rectangle. This fact is illustrated in Fig. 4.3. The thick lines represent the precise boundary of the regions described by (4.8), (4.9), and (4.10). The thin lines denote the approximate upper and lower bounds determined from (4.11) and (4.12). Epipolar spaces are nonuniform in area, increasing in size with increasing values of u and v.

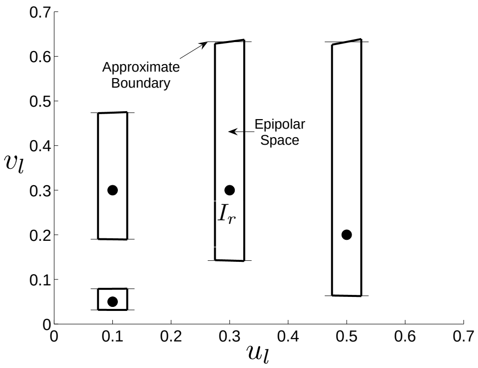


Figure 4.3: Various epipolar spaces. The large dots represent points from the right image. Each point is enclosed in its attendant epipolar space in the left image. The thick lines represent the precise boundaries described by (4.8), (4.9), and (4.10). The thin lines denote the approximate upper and lower bounds determined from (4.11) and (4.12).

4.3 Optimal Sampling Theorems

Epipolar spaces eliminate the need for calibration at the expense of an increased search space. To mitigate this increase, a sampling strategy can be applied. The optimal strategy is defined as the one that minimizes the average number of points per epipolar space. Though sampling is inherently a discrete process, it can be modeled continuously. The purpose of this section is to formulate and derive these optimal sampling strategies using continuous variables and functions. Working in a continuous domain simplifies the analysis, allowing the use of powerful mathematical tools that are either unavailable or extremely cumbersome in a discrete framework.

The derivations in the following subsections are not restricted to the specific epipolar spaces described by (4.8), (4.11), and (4.12), but instead, are generalized to epipolar spaces of a variety of shapes and sizes. Section 4.3.1 considers the optimal sampling strategies for a one-dimensional image interval. Analysis in one-dimension simplifies the mathematics and provides a sound framework for the two-dimensional adaptation. Section 4.3.2 extends the optimal one-dimensional sampling strategies to two dimensions.

4.3.1 One-Dimensional

Consider any one-dimensional interval I defined by $[a, b]$. For each point u in I the matching point must lie within the corresponding epipolar interval. The upper and lower extents of these epipolar intervals can be expressed as functions of u. Let $i\,(u)$ represent the epipolar interval associated with the point u; and let $i_b\,(u)$ and $i_a\,(u)$ indicate the upper and lower bounds of $i\,(u)$, respectively. The goal of an optimal sampling strategy is to place N points in the interval I in such a fashion as to minimize the average number of candidate matches for each of the N points. A candidate match for point u_j is defined as any point u_k that lies within the epipolar interval $i\,(u_j)$.

A discrete sampling strategy can be modeled as a continuous sampling function. A sampling function is defined as any increasing, invertible function $\gamma\,(u)$. For a given sampling strategy $\gamma\,(u)$ and set of epipolar intervals $i\,(u)$, the average epipolar length $E\,(\gamma; i, I)$ over I can be expressed as

$$E\,(\gamma; i, I) = \frac{1}{b-a} \int\limits_a^b e\,[\gamma; i\,(u)]\, \gamma'\,(u)\, du, \tag{4.14}$$

where

$$e\left[\gamma; i\left(u\right)\right] = \int\limits_{i_a(u)}^{i_b(u)} \gamma'\left(\tilde{u}\right) d\tilde{u} \tag{4.15}$$

is the length of the epipolar interval associated with the point u after transformation. The optimal sampling function is the function $\gamma\left(u\right)$ that minimizes (4.14) subject to the constraint of length preservation:

$$\Lambda\left(\gamma; I\right) = \int\limits_{a}^{b} \gamma'\left(u\right) du = b - a. \tag{4.16}$$

Since all the followings proofs depend solely on the derivative of $\gamma\left(u\right)$, the choice of the constant term in $\gamma\left(u\right)$ is arbitrary. Therefore, let any $\gamma\left(u\right)$ satisfying (4.16) map $[a, b]$ onto itself, i.e. $\gamma\left(a\right) = a$, $\gamma\left(b\right) = b$. Mapping to the same interval removes ambiguity and conceptually simplifies future derivations without reducing the generality of the results.

We begin by considering uniform epipolar intervals $\tilde{i}\left(u\right)$ defined by $[\max\left\{u-c, a\right\}, \min\left\{b, u+c\right\}]$, where $c \geq 0$. The min and max functions clip the epipolar intervals that extend outside of $[a, b]$. In the modeling of practical applications it is reasonable to assume that $b-a \gg c$. This stipulation eliminates the need for clipping, simplifying the interval to $[u-c, u+c]$. Furthermore, under this assumption (4.15) can be well approximated (for uniform epipolar intervals) as follows:

$$\begin{aligned} e\left[\gamma; \tilde{i}\left(u\right)\right] &\approx 2c\gamma'\left(u\right) \\ &= \tilde{e}\left(\gamma; c\right). \end{aligned} \tag{4.17}$$

Inserting this result into (4.14) produces

$$\begin{aligned} E\left(\gamma;\tilde{i},I\right) &\approx \frac{2c}{b-a}\int\limits_a^b \left[\gamma'\left(u\right)\right]^2 du \\ &= \tilde{E}\left(\gamma;c,I\right). \end{aligned} \tag{4.18}$$

The next portion of this section provides a sequence of theorems and their proofs that will, in combination, provide a framework for obtaining optimal sampling schemes for a given set of epipolar intervals. A discussion of their significance immediately follows their presentation.

Theorem 4.3.1. If $\gamma\left(u\right)\in \mathrm{C}^2\left[a,b\right]$ then $\gamma\left(u\right)=u$ is the unique minimizer of the objective functional given in (4.18) subject to the constraint posed in (4.16).

Proof. The Euler-Lagrange equation provides the first order necessary condition for a local minimizer of (4.18):

$$\frac{\partial L}{\partial \gamma}-\frac{d}{dx}\left(\frac{\partial L}{\partial \gamma'}\right)=0, \tag{4.19}$$

where $L\left(u,\gamma,\gamma'\right)$ is the Lagrangian [85]. Inserting $L\left(u,\gamma,\gamma'\right)=\left[\gamma'\left(u\right)\right]^2$ into (4.19) produces

$$\gamma''\left(u\right)=0. \tag{4.20}$$

Integrating this result yields $\gamma\left(u\right)=\beta u+\alpha$. The constraint prescribed in (4.16) forces $\beta=1$. Since it was previously stipulated that all sampling functions map I onto itself, it follows that $\alpha=0$. In order to guarantee that this single critical function is a minimizer, it is necessary to examine the second variation $Q\left(\gamma;v\right)$:

$$Q\left(\gamma;v\right)=\int\limits_a^b \left[Av^2+2Bvv'+C\left(v'\right)^2\right]du, \tag{4.21}$$

where

$$A = \frac{\partial^2 L}{\partial\gamma\partial\gamma}, \quad B = \frac{\partial^2 L}{\partial\gamma\partial\gamma'}, \quad C = \frac{\partial^2 L}{\partial\gamma'\partial\gamma'},$$

and $v(u) \in \mathrm{C}^2[a,b]$ is any arbitrary function with $v(a) = 0 = v(b)$. If $Q(\gamma; v) > 0$ for all v not identically zero, then $Q(\gamma; v)$ is positive definite and $\gamma(u)$ is a strict local minimum [85]. Inserting $L(u, \gamma, \gamma') = [\gamma'(u)]^2$ into the (4.21) produces

$$Q(\gamma; v) = \int_a^b 2(v')^2 \, du. \tag{4.22}$$

This quantity vanishes only if v is a constant. Since $v(a) = 0$, it follows that $v \equiv 0$. Therefore, (4.21) is positive definite and $\gamma(u) = u$ is the unique minimizer of (4.18). □

Theorem 4.3.2. If $\gamma(u) = \Gamma(u)$ is the sampling function that minimizes the functional $E(\gamma; i, I)$ subject to (4.16) and $\chi(u)$ is another arbitrary sampling function satisfying (4.16), then $\gamma(u) = \Gamma_\chi(u) = \Gamma(\chi(u))$ minimizes $E(\gamma; i_\chi, I)$ subject to (4.16), where $i_\chi(u)$ is the interval with upper bound $i_{b\chi}(u) = \chi^{-1}(i_b[\chi(u)])$ and lower bound $i_{a\chi}(u) = \chi^{-1}(i_a[\chi(u)])$.

Proof. The following equations use the variable substitution $x = \chi^{-1}(u)$.

$$\begin{aligned} E(\Gamma; i, I) &= \frac{1}{b-a} \int_a^b e[\Gamma; i(u)]\, \Gamma'(u)\, du \\ &= \frac{1}{b-a} \int_a^b e(\Gamma; i[\chi(x)])\, \Gamma'[\chi(x)]\, \chi'(x)\, dx \\ &= \frac{1}{b-a} \int_a^b e[\Gamma_\chi; i_\chi(x)]\, \Gamma'_\chi(x)\, dx \\ &= E(\Gamma_\chi; i_\chi, I) \end{aligned} \tag{4.23}$$

$$\begin{aligned} e(\Gamma; i[\chi(x)]) &= \int_{i_a(\chi(x))}^{i_b(\chi(x))} \Gamma'(\tilde{u})\, d\tilde{u} \\ &= \int_{\chi^{-1}(i_a[\chi(x)])}^{\chi^{-1}(i_b[\chi(x)])} \Gamma'[\chi(\tilde{x})]\, \chi'(\tilde{x})\, d\tilde{x} \\ &= \int_{i_{a\chi}(x)}^{i_{b\chi}(x)} \Gamma'_\chi(\tilde{x})\, d\tilde{x} \\ &= e[\Gamma_\chi; i_\chi(x)] \end{aligned} \tag{4.24}$$

$$\begin{aligned} \Lambda(\Gamma_\chi; I) &= \int_a^b \Gamma'[\chi(u)]\, \chi'(u)\, du \\ &= \int_a^b \Gamma'(x)\, dx \\ &= \Lambda(\Gamma; I) = b - a. \end{aligned} \tag{4.25}$$

□

From the above proofs we can draw several important conclusions. First, Theorem 4.3.1 demonstrates that uniform sampling becomes optimal for uniform epipolar intervals when the total interval length is large with respect to the lengths of the uniform epipolar intervals. This caveat of relative length is necessary to mitigate the effects of the nonlinear clipping needed to ensure that the epipolar intervals never extend outside I. Second, Theorem 4.3.2 provides a means for creating other sets of epipolar intervals and their attendant optimal sampling functions by warping uniform epipolar intervals and their corresponding optimal sampling function $\Gamma(u) = u$. This immediately suggests the following corollary: *if a sampling function $\gamma(u)$ warps intervals of nonuniform length into intervals of uniform length, it is the optimal sampling function with respect to those nonuniform intervals.* This corollary becomes readily apparent by reversing the proof of Theorem 4.3.2 with $\Gamma(u) = u$.

4.3.2 Two-Dimensional

Consider any two-dimensional, connected region R. For each point $\boldsymbol{u}$ in R the matching point must lie within the corresponding epipolar space $r(\boldsymbol{u})$. The goal of a two-dimensional optimal sampling strategy is to place N points in the region R in such a fashion as to minimize the average number of candidate matches for each of the N points. A candidate match for point $\boldsymbol{u}_j$ is defined as any point $\boldsymbol{u}_k$ that lies within the epipolar space $r(\boldsymbol{u}_j)$.

A two-dimensional continuous sampling function is defined as any invertible function $\boldsymbol{\gamma}(\boldsymbol{u}) \doteq [\gamma_u(\boldsymbol{u}), \gamma_v(\boldsymbol{u})]^t$ that maps $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ and has a Jacobian matrix with a determinant that is positive everywhere. For a given sampling function $\boldsymbol{\gamma}(\boldsymbol{u})$ the average area of the epipolar spaces $r(\boldsymbol{u})$ over the region R is defined as

$$E(\boldsymbol{\gamma}; r, R) = \frac{1}{A_R} \iint_R e[\boldsymbol{\gamma}; r(\boldsymbol{u})] |\boldsymbol{\gamma}'(\boldsymbol{u})| d\boldsymbol{u}, \tag{4.26}$$

where

$$e[\boldsymbol{\gamma}; r(\boldsymbol{u})] = \iint_{r(u)} |\boldsymbol{\gamma}'(\tilde{\boldsymbol{u}})| \, d\tilde{\boldsymbol{u}} \tag{4.27}$$

is the area of the epipolar space associated with the point $\boldsymbol{u}$ after transformation, $\boldsymbol{\gamma}'(\boldsymbol{u})$ is the Jacobian matrix, and $A_R = \iint_R d\boldsymbol{u}$. An optimal sampling scheme is a function $\boldsymbol{\gamma}(\boldsymbol{u})$ that minimizes $E(\boldsymbol{\gamma}; r, R)$ subject to the constraint of area preservation:

$$\Lambda(\boldsymbol{\gamma}; R) = \iint_R |\boldsymbol{\gamma}'(\boldsymbol{u})| \, d\boldsymbol{u} = \iint_{A_R} d\boldsymbol{u} = A_R. \tag{4.28}$$

We again begin by considering uniform rectangular epipolar spaces $\tilde{r}(\boldsymbol{u})$ defined by $\tilde{u} \in \tilde{i}_u(u)$ and $\tilde{v} \in \tilde{i}_v(v)$, where the intervals $\tilde{i}_u(u)$ and $\tilde{i}_v(v)$ are given by $[u - c_u, u + c_u]$ and $[v - c_v, v + c_v]$, respectively, with $c_u \geq 0$ and $c_v \geq 0$. As in the one-dimensional case, it is assumed that $A_R \gg c_u c_v$, eliminating the need to clip the epipolar spaces that extend outside of R. Furthermore, since the epipolar spaces are separable and not spatially variant, it is reasonable to restrict the space of optimal sampling functions to the space of separable functions, i.e. $\boldsymbol{\gamma}(\boldsymbol{u}) \doteq [\gamma_u(u), \gamma_v(v)]^t$. With these restrictions (4.26) can be well approximated as follows:

$$\begin{aligned} e(\boldsymbol{\gamma}; \tilde{r}) &\approx 4 c_u c_v \gamma_u'(u) \gamma_v'(v) \\ &= \tilde{e}(\boldsymbol{\gamma}; \boldsymbol{c}), \end{aligned} \tag{4.29}$$

where $\boldsymbol{c} = [c_u, c_v]^t$. Assuming separability, the insertion of (4.29) into (4.26) produces

$$\begin{aligned} E(\boldsymbol{\gamma}; \tilde{r}, R) &\approx \frac{4c_u c_v}{A_R} \iint_R [\gamma_u'(u)\,\gamma_v'(v)]^2\, dudv \\ &= \tilde{E}(\boldsymbol{\gamma}; \boldsymbol{c}, R) \end{aligned} \tag{4.30}$$

The following theorems and proofs form the foundation for constructing optimal two-dimensional sampling strategies.

Theorem 4.3.3. Let R be a rectangular image plane with $u \in I_u$ and $v \in I_v$, where I_u and I_v are the intervals defined by $[a_u, b_u]$ and $[a_v, b_v]$, respectively. If $\gamma_u(u) \in \mathrm{C}^2[a_u, b_u]$ and $\gamma_v(v) \in \mathrm{C}^2[a_v, b_v]$ then $\gamma_u(u) = \beta_u u$ and $\gamma_v(v) = \beta_v v$, where $\beta_u \beta_v = 1$, minimize the objective functional given in (4.30), subject to the constraint posed in (4.28).

Proof.

$$\begin{aligned} \tilde{E}(\boldsymbol{\gamma}; \boldsymbol{c}, R) &= \frac{4c_u c_v}{A_R} \iint_R [\gamma_u'(u)\,\gamma_v'(v)]^2\, dudv \\ &= \frac{2c_u}{b_u - a_u} \int\limits_{a_u}^{b_u} [\gamma_u'(u)]^2\, du \cdot \\ &\qquad \frac{2c_v}{b_v - a_v} \int\limits_{a_v}^{b_v} [\gamma_v'(v)]^2\, dv \\ &= \tilde{E}(\gamma_u; c_u, I_u)\, \tilde{E}(\gamma_v; c_v, I_v) \end{aligned} \tag{4.31}$$

Equation (4.31) obtains its minimum when $\tilde{E}(\gamma_u; c_u, I_u)$ and $\tilde{E}(\gamma_v; c_v, I_v)$ obtain their respective minimums. From the proof of Theorem 4.3.1 we know this occurs when $\gamma_u = \beta_u u + \alpha_u$ and $\gamma_v = \beta_v v + \alpha_v$. The constraint in (4.28) forces $\beta_u \beta_v = 1$. Since it was stipulated that any one-dimensional sampling function maps I onto itself, it follows that $\alpha_u = 0$ and $\alpha_v = 0$. $\square$

Theorem 4.3.4. If $\boldsymbol{\gamma}(\boldsymbol{u}) = \boldsymbol{\Gamma}(\boldsymbol{u})$ is a sampling function that minimizes the functional $E(\boldsymbol{\gamma}; r, R)$ subject to $\Lambda(\boldsymbol{\gamma}; R) = A_R$ and $\boldsymbol{\chi}(\boldsymbol{u})$ is some other sampling function satisfying $\Lambda(\boldsymbol{\chi}^{-1}; R) = A_R$, then $\boldsymbol{\gamma}(\boldsymbol{u}) = \boldsymbol{\Gamma}_\chi(\boldsymbol{u}) = \boldsymbol{\Gamma}(\boldsymbol{\chi}(\boldsymbol{u}))$ minimizes $E(\boldsymbol{\gamma}; r_\chi, R_\chi)$ subject to $\Lambda(\boldsymbol{\gamma}; R_\chi) = A_{R_\chi}$, where $R_\chi = \boldsymbol{\chi}^{-1}(R)$ and $r_\chi(\boldsymbol{u}) = \boldsymbol{\chi}^{-1}(r[\boldsymbol{\chi}(\boldsymbol{u})])$.

Proof. The following equations use the variable substitution $\boldsymbol{x} = \boldsymbol{\chi}^{-1}(\boldsymbol{u})$.

By definition, $\Lambda(\boldsymbol{\chi}^{-1}; R) = A_R$ and $\Lambda(\boldsymbol{\Gamma}; R) = A_R$, consequently

$$\begin{aligned} A_R &= \iint_R \left| \left(\boldsymbol{\chi}^{-1}\right)'(\boldsymbol{u}) \right| d\boldsymbol{u} \\ &= \iint_{R_\chi} d\boldsymbol{x} = A_{R_\chi} \end{aligned} \tag{4.32}$$

$$\begin{aligned} A_{R_\chi} = A_R &= \iint_R \left| \boldsymbol{\Gamma}'(\boldsymbol{u}) \right| d\boldsymbol{u} \\ &= \iint_{R_\chi} \left| \boldsymbol{\Gamma}'[\boldsymbol{\chi}(\boldsymbol{x})] \right| \left| \boldsymbol{\chi}'(\boldsymbol{x}) \right| d\boldsymbol{x} \\ &= \iint_{R_\chi} \left| \boldsymbol{\Gamma}'_\chi(\boldsymbol{x}) \right| d\boldsymbol{x} \\ &= \Lambda(\boldsymbol{\Gamma}_\chi; R_\chi). \end{aligned} \tag{4.33}$$

Using these results and the aforementioned variable substitution yields:

$$\begin{aligned}
E\left(\boldsymbol{\Gamma}; r, R\right) &= \frac{1}{A_R} \iint\limits_{R} e\left[\boldsymbol{\Gamma}; r\left(\boldsymbol{u}\right)\right] \left|\boldsymbol{\Gamma}'\left(\boldsymbol{u}\right)\right| d\boldsymbol{u} \\
&= \frac{1}{A_R} \iint\limits_{\chi^{-1}(R)} e\left(\boldsymbol{\Gamma}; r\left[\boldsymbol{\chi}\left(\boldsymbol{x}\right)\right]\right) \cdot \\
&\qquad \left|\boldsymbol{\Gamma}'\left[\boldsymbol{\chi}\left(\boldsymbol{x}\right)\right]\right| \left|\boldsymbol{\chi}'\left(\boldsymbol{x}\right)\right| d\boldsymbol{x} \\
&= \frac{1}{A_{R_\chi}} \iint\limits_{R_\chi} e\left[\boldsymbol{\Gamma}_\chi; r_\chi\left(\boldsymbol{x}\right)\right] \left|\boldsymbol{\Gamma}'_\chi\left(\boldsymbol{x}\right)\right| d\boldsymbol{x} \\
&= E\left(\boldsymbol{\Gamma}_\chi; r_\chi, R_\chi\right) \qquad (4.34)
\end{aligned}$$

$$\begin{aligned}
e\left(\boldsymbol{\Gamma}; r\left[\boldsymbol{\chi}\left(\boldsymbol{x}\right)\right]\right) &= \iint\limits_{r(\chi(x))} \left|\boldsymbol{\Gamma}'\left(\tilde{\boldsymbol{u}}\right)\right| d\tilde{\boldsymbol{u}} \\
&= \iint\limits_{\chi^{-1}(r[\chi(x)])} \left|\boldsymbol{\Gamma}'\left[\boldsymbol{\chi}\left(\tilde{\boldsymbol{x}}\right)\right]\right| \left|\boldsymbol{\chi}'\left(\tilde{\boldsymbol{x}}\right)\right| d\tilde{\boldsymbol{x}} \\
&= \iint\limits_{r_\chi(x)} \left|\boldsymbol{\Gamma}'_\chi\left(\tilde{\boldsymbol{x}}\right)\right| d\tilde{\boldsymbol{x}} \\
&= e\left[\boldsymbol{\Gamma}_\chi; r_\chi\left(x\right)\right] \qquad (4.35)
\end{aligned}$$

□

For practical purposes it is important to note that though Theorem 4.3.3 was presented in the context of a rectangular image plane R, the result still holds for any arbitrarily shaped region R with the following caveat: the total area clipped from the epipolar spaces that extend outside of R must be sufficiently small compared to the total area of R. In general, the size and shape of the regions R and R_χ can be arbitrarily chosen to fit a specific circumstance, again, with the caveat that the total clipped epipolar area be relatively small.

The significance of the previous proofs is now addressed. Theorem 4.3.3 states that uniform rectangular sampling is optimal for uniform rectangular epipolar spaces when the area of R is large with respect to the individual epipolar areas, i.e. as the amount of clipping becomes negligible. This holds for uniform rectangular epipolar spaces of any aspect ratio. It also holds for rectangular sampling of any aspect ratio; that is, the sampling rate in each dimension need not agree. Combining this result with Theorem 4.3.4 demonstrates that any area preserving sampling function (i.e., it satisfies (4.28)) that warps nonuniformly sized epipolar spaces into rectangles of uniform size is the minimizing function associated with those nonuniform spaces. In fact, the nonuniform spaces do not have to be warped into rectangles. They need only be mapped into regions of uniform area. To see this, first consider the area preserving sampling function $\boldsymbol{\chi}(\boldsymbol{u})$ that warps the uniform rectangular epipolar spaces $\tilde{r}(\boldsymbol{u})$ of area C_e into the arbitrarily shaped regions $\tilde{r}_\chi(\boldsymbol{u})$ of identical area C_e. Substituting $\tilde{r}(\boldsymbol{u})$ and the optimal sampling function $\gamma(\boldsymbol{u}) = \boldsymbol{u}$ into (4.26) and then using the variable substitution $\boldsymbol{x} = \boldsymbol{\chi}^{-1}(\boldsymbol{u})$ yields

$$\begin{aligned} E(\boldsymbol{u}; \tilde{r}, R) &= \frac{1}{A_R} \iint_R C_e \, d\boldsymbol{u} \\ &= \frac{1}{A_{R_\chi}} \iint_{R_\chi} C_e \left|\boldsymbol{\chi}'(\boldsymbol{x})\right| d\boldsymbol{x} \\ &= E(\boldsymbol{\chi}; \tilde{r}_\chi, R_\chi). \end{aligned} \tag{4.36}$$

In summary, *any area preserving sampling function* $\boldsymbol{\gamma}(\boldsymbol{u})$ *that warps epipolar* *spaces into regions of uniform area is optimal with respect to their average epipolar area.*

4.4 Optimal Sampling Applications

This section uses the results from previous sections to formulate the optimal sampling schemes for the epipolar spaces defined in (4.8), (4.11), and (4.12). First, the optimal scheme is derived independently for each dimension. Then these solutions are integrated to create the optimal two-dimensional sampling scheme.

4.4.1 One-Dimensional

The epipolar spaces delineated in (4.8), (4.11), and (4.12) are identical in both images and symmetric across image quadrants, therefore we need only consider the positive coordinates of a single image plane. For the point (u, v), the corresponding bounds for a one-dimensional epipolar interval $i_1(u)$ along the horizontal u dimension are given by (4.8). These boundaries form intervals of uniform length $2D$ (ignoring clipping). As shown in Section 4.3.1, the optimal sampling scheme over the interval I for epipolar intervals of uniform length is

$$\gamma_1(u) = u. \tag{4.37}$$

The average epipolar length for these uniform epipolar intervals under uniform sampling is

$$E(\gamma_1; i_1, I) = 2D - \frac{D^2}{b-a}. \tag{4.38}$$

This converges to $2D$ as $b-a$ becomes large with respect to D.

In the vertical v dimension the boundaries of the epipolar interval $i_2(v)$ are given by (4.11) and (4.12). The resulting intervals $[v/c(u), vc(u)]$ are nonuniform in length. The function $\gamma(v)$ that minimizes their average length on I must warp them into intervals of uniform length and satisfy (4.16). This optimal function is

$$\gamma_2(v) = \beta \ln v, \tag{4.39}$$

where $\beta = (b-a)/(\ln b - \ln a)$. (For simplicity, we have set the constant term of $\gamma_2(v)$ to zero. With this choice the sampling function does not map I onto itself.) The average epipolar length under this logarithmic transformation is

$$E(\gamma_v; i_2, I) = 2\beta \ln c - \frac{[\beta \ln c]^2}{b-a}. \tag{4.40}$$

Since (4.39) has a singularity at zero, the interval $[a, b]$ must not contain zero.

If instead the interval were sampled uniformly, the average epipolar length would be

$$E(\gamma_1; i_2, I) = \frac{a^2(1-c) + b^2(1-1/c)}{b-a}. \tag{4.41}$$

Fig. 4.4 plots the ratio $E(\gamma_1; i_2, I)/E(\gamma_2, i_2, I)$. Since the value of the variable c given in (4.13) is a function of u, this ratio is plotted over a range positive u values. For small values of u, the logarithmic sampling function $\gamma_2(v)$ produces epipolar lengths that are, on average, less than half the size of those produced by the uniform sampling function $\gamma_1(v)$.

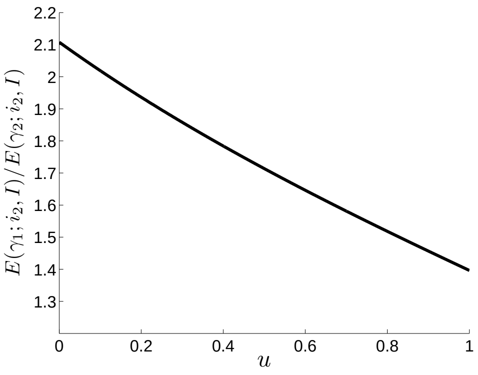


Figure 4.4: Ratio of the mean lengths $E(\gamma_1; i_2, I)/E(\gamma_2; i_2, I)$ of the epipolar intervals delineated in (4.11) and (4.12) under uniform and logarithmic sampling. The vertical interval extends from $a = 0.017$ to $b = 1$. Since the value of the variable $c(u)$ described in (4.13) is a function of u, the ratio is plotted over a range of positive u values. The determination of $c(u)$ uses the following parameters: $\theta_M = \pi/3$ and $f = 1$.

For additional insight, consider a third sampling method that, like logarithmic sampling, also tends to concentrate more samples near a:

$$\gamma_3(v) = -abv^{-1}, \tag{4.42}$$

The commensurate average epipolar length is

$$E(\gamma_3; i_2, I) = \frac{a^2(1-c) + b^2(1-1/c)}{b-a}. \tag{4.43}$$

Interestingly, this is identical to the average length for uniform sampling.

4.4.2 Two-Dimensional

In the previous subsection each dimension was optimized independently. Unfortunately, the vertical bounds of the epipolar spaces described in (4.11) and (4.12) depend on both u and v: they are nonseparable. To produce regions of equal area (and, therefore, minimize the average epipolar area) we must compensate for this dependence by normalizing (4.39). The following function produces spaces of approximately equal area:

$$\boldsymbol{\gamma}(u,v) = \left[\beta_u u,\ \frac{\beta_v \ln v}{\ln c(u)}\right]^t, \tag{4.44}$$

where β_u and β_v are constants chosen such that their product satisfies the constraint in (4.28). The following calculation of the area of an arbitrary epipolar space $r(\boldsymbol{u})$ (that is not clipped) after transformation confirms they have constant area:

$$\begin{aligned} e(\boldsymbol{\gamma}; r) &= \int_{u-D}^{u+D} \int_{v/c(u)}^{vc(u)} \frac{\beta_u \beta_v}{v \ln[c(\tilde{u})]} d\tilde{u} d\tilde{v} \\ &= \int_{u-D}^{u+D} 2\beta_u \beta_v \frac{\ln c(u)}{\ln c(\tilde{u})} d\tilde{u} \\ &\approx 4D\beta_u\beta_v. \end{aligned} \tag{4.45}$$

This follows from the fact that for $\tilde{u} \in [u-D, u+D]$ the function $\frac{\ln c_v(u)}{\ln c_v(\tilde{u})}$ is approximately linear with a mean value of one. Fig. 4.5 illustrates the average area of each epipolar space following transformation. As expected this average remains constant over the majority of the image plane. The degradation seen in the upper right corner is caused by clipping. The epipolar spaces in this region have relatively large areas extending outside the image boundary.

In Section 4.2 the epipolar spaces described in (4.8), (4.11), and (4.12) were depicted as Fig. 4.3. The transformations of these spaces using (4.44) are

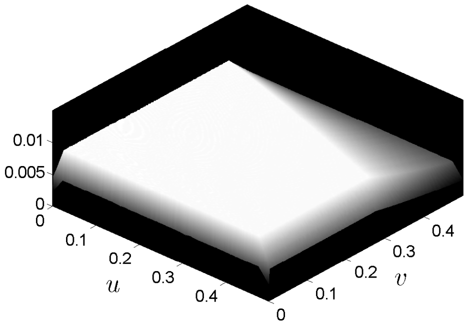


Figure 4.5: Area per epipolar space. Each area is normalized by the total area of the image plane. The normalized area remains constant over the majority of the image plane. The degradation seen in the upper right corner is caused by clipping.

shown in Fig. 4.6. Though the warped boundaries are not identical in shape, they are almost perfectly uniform in area.

As mentioned previously, the constants β_u and β_v in (4.44) are only restricted in the sense that their product must satisfy (4.28). This is a consequence of the fact that the average epipolar area is a function of the determinant of the Jacobian of $\boldsymbol{\gamma}(\boldsymbol{u})$, and not $\boldsymbol{\gamma}(\boldsymbol{u})$ itself. That is, an infinite number of functions may have Jacobian matrices with the identical determinant. Fig. 4.7 illustrates the optimal sampling strategies when β_u is set to two different, arbitrarily chosen values: $\beta_u = 1$ and $\beta_u = 4$.

It may be beneficial to put these results into context by reviewing the meaning of an epipolar space, its importance in stereo registration, and its connection to the optimal sampling strategies. If a point in space projects onto the left image plane, its projection onto the right image plane is restricted to

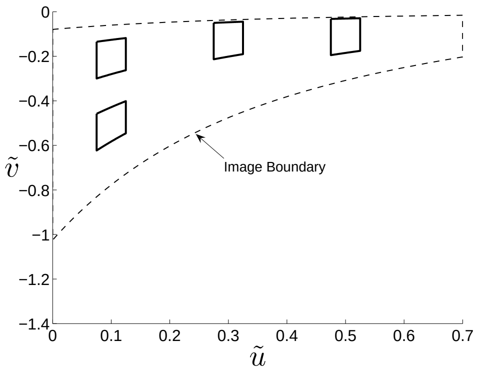


Figure 4.6: Epipolar spaces from Fig. 4.3 after optimal sampling. The epipolar spaces in Fig. 4.3 were warped using the transformation in (4.44) with $\beta_u = 1$. They are almost perfectly uniform in area.

a region called an epipolar space. From a point of view of a stereo registration algorithm, only the pixels in the epipolar space must be examined to identify the matching point. The complexity of this search is proportional to the number of pixels in the epipolar space. Consider the epipolar spaces shown in Fig. 4.3. If pixels were to uniformly cover this space, then the larger epipolar spaces would require a greater search time than the smaller spaces. However, the time to search a particular epipolar space is not as important as the time it takes to search them all, which is what must be done for dense stereo matching. Now consider if the epipolar spaces in Fig. 4.3 were placed on a sampling grid like those depicted Fig. 4.7. Were we to count the number of samples appearing in each epipolar space, we would find that they are identical; and this number would be the same for both sampling patterns in Fig. 4.7. More importantly, if a fixed number of samples were placed within the defined image boundary

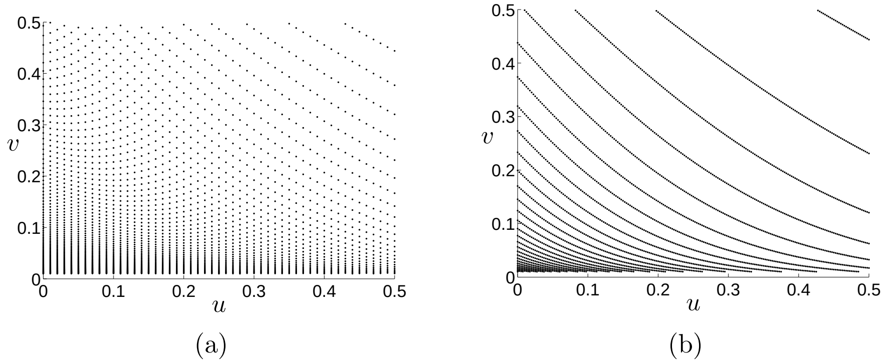


Figure 4.7: Example of optimal sampling scheme defined in (4.44) for two different values of β_u. Given β_u, the value of β_v is determined by (4.28). Additional parameters were also assigned as follows: $f=1$, $\theta_M=\pi/3$, $a=0.001$, and $b=0.5$. Fig. 4.7(a) shows the optimal sampling scheme with $\beta_u=1$. Fig. 4.7(b) shows the optimal sampling scheme with $\beta_u=4$.

and then the average number of points per epipolar space were computed, no placement strategy would produce an average less than that of those shown in Fig. 4.7. As an example, Table 4.1 provides a quantified comparison of the average number of samples per epipolar space for both uniform sampling and the optimal sampling strategy described in (4.44). The data were generated for various rectangular image sizes and maximum rotation angles θ_M.

4.5 Binocular Active Vision

This section discusses how previous results are applicable to active binocular vision systems. We construct a foveated tessellation as a means for implementing the optimal nonuniform sampling strategy. Then, we demonstrate how this tessellation could improve the accuracy and performance of

a_v	θ_M	$E(\boldsymbol{\gamma}_1; r, R) / E(\boldsymbol{\gamma}_2; r, R)$
0.1	$\pi/4$	1.74
0.1	$\pi/3$	2.06
0.01	$\pi/4$	2.58
0.01	$\pi/3$	3.12
0.001	$\pi/4$	3.47
0.001	$\pi/3$	4.22

Table 4.1: Ratio of samples per epipolar space for uniform and optimal sampling. Equivalently, this table presents the ratio of the average area of an epipolar space before and after transformation by the sampling function posed in (4.44). The leftmost, rightmost, and upper bounds for the rectangular image are $a_u = 0$, $b_u = 0.5$, and $b_v = 0.5$, respectively. The lower bound is specified in the table.

a dense stereo matching algorithm that is suitable for use with uncalibrated ABVS that employ nonuniform sampling.

4.5.1 Antialiasing with Tessellations

Since using specialized hardware or optics is often prohibitive, we will assume that all foveated images are created by using software to nonuniformly sample typical uniformly sampled images. In order to prevent aliasing, the images must be filtered prior to sampling. Since nonuniform sampling is inherently space-variant, so are the antialiasing filters. First consider the continuous two-dimensional function $f(\boldsymbol{u})$ which represents a continuous image. Before sampling $f(\boldsymbol{u})$ at $\boldsymbol{u}_{\boldsymbol{n}} = \boldsymbol{n}$, the image should be bandlimited to $[-\pi, \pi]$ in each dimension. Using the ideal filter to allow for perfect reconstruction, this is done as follows:

$$f_{\Phi}(\boldsymbol{u}) = \iint\limits_{-\infty}^{\infty} f(\boldsymbol{\xi})\,\Phi(\boldsymbol{u}-\boldsymbol{\xi})\,d\boldsymbol{u}, \tag{4.46}$$

where

$$\Phi(\boldsymbol{u}) = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \exp\left(j\boldsymbol{\omega}^t\boldsymbol{u}\right) d\boldsymbol{\omega}. \tag{4.47}$$

Now consider the continuous sampling function $\boldsymbol{\gamma}(\boldsymbol{u})$ as defined in Section 4.3.2. To prevent aliasing when $f(\boldsymbol{u})$ is sampled at $\boldsymbol{u}_n = \boldsymbol{\gamma}^{-1}(\boldsymbol{n})$, the image $f(\boldsymbol{u})$ must be locally bandlimited as follows [28, 134]:

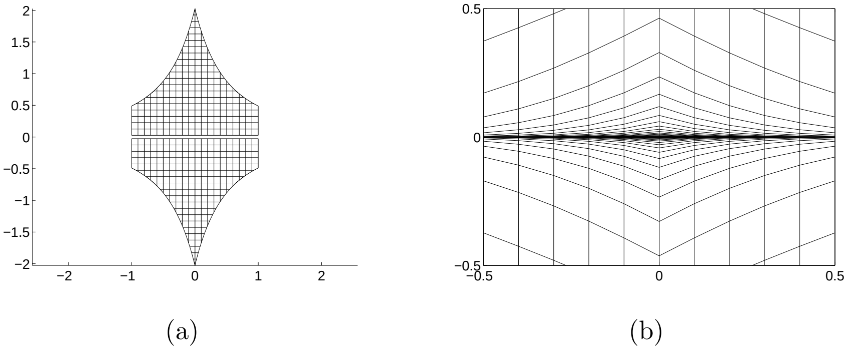
$$f_{\Phi_\gamma}(\boldsymbol{u}) = \iint\limits_{-\infty}^{\infty} f(\boldsymbol{\xi})\,\Phi_\gamma(\boldsymbol{u},\boldsymbol{\xi})\,d\boldsymbol{\xi}, \tag{4.48}$$

where

$$\Phi_\gamma(\boldsymbol{u},\boldsymbol{\xi}) = \Phi\left[\boldsymbol{\gamma}(\boldsymbol{u}) - \boldsymbol{\gamma}(\boldsymbol{\xi})\right]\boldsymbol{\gamma}'(\boldsymbol{\xi}). \tag{4.49}$$

The function $\Phi(\boldsymbol{u},\boldsymbol{\xi})$ can be considered a space-variant antialiasing filter that conforms to the local spacing of the samples. As the samples become sparser, the filter $\Phi(\boldsymbol{u},\boldsymbol{\xi})$ becomes increasingly low-pass. Theoretically, $f_{\Phi_\gamma}(\boldsymbol{u})$ could be perfectly reconstructed from its samples $f_{\Phi_\gamma}(\boldsymbol{u}_n)$ [28]. Filtering an image using (4.49) with $\boldsymbol{\gamma}(u)$ equal to the optimal sampling function presented in (4.44) would create a foveated image appropriate for sampling with the previously defined optimal sampling strategy.

For reasons analogous to those which discourage the use of the sinc function as an antialiasing filter, the ideal antialiasing function $\Phi(\boldsymbol{u},\boldsymbol{\xi})$ as defined in (4.49) is seldom preferred. Fortunately, there are a myriad of means for approximating its effects. A popular and intuitive approach uses super-pixels. Consider the unform grid in Fig. 4.8(a). This grid represents the uniform sampling of the warped image space resulting from the application of the optimal transformation in (4.44) to a square image plane. Fig. 4.8(b) illustrates the tessellation resulting from projecting the uniform grid in Fig. 4.8(a) back into the original square image plane. Each enclosed area in the tessellation is a super-pixel. Foveating a typical uniformly sampled image requires assigning all the uniform pixels within each super-pixel their average value. In the warped space each uniform pixel is assigned the average of its concomitant super-pixel.



(a) (b)

Figure 4.8: Tessellations for optimal epipolar sampling. Fig. 4.8(a) illustrates the warped image space resulting from the application of the optimal transformation in (4.44) to a square image plane. Fig. 4.8(b) shows the projection of the uniform grid in 4.8(a) back into the original image plane.

4.5.2 Application to Active Vision

We now discuss how an off-the-shelf stereo registration algorithm could use the results in this dissertation to improve its accuracy and efficiency when participating in an ABVS. First, assume that the geometric configuration of the ABVS conforms to the geometric stipulations in Section 4.2, i.e. the only movable parts effecting epipolar geometry are the camera rotation angles and their range is restricted by (4.7). With this restriction and the restriction on horizontal disparity described in (4.8), an optimal tessellation like that shown in Fig. 4.8(b) can be constructed. Any off-the-shelf uncalibrated (i.e. it does not require knowledge of the geometry) stereo registration algorithm designed for foveated images can use this tessellation to improve its performance and accuracy.

Currently, there is a paucity of dense registration algorithms suitable for foveated images. In fact, we are aware of only one algorithm that is appropriate for use with a uncalibrated ABVS employing an arbitrary sampling strategy. In [13], Bernardino uses a Bayesian formulation to determine the conditional probability that at the point $\boldsymbol{u}$ the discrete disparity vector is $\boldsymbol{d}_n(\boldsymbol{u})$. This probability is conditioned on the intensity values $I_l(\boldsymbol{u})$ and $I_r(\boldsymbol{u}+\boldsymbol{d}_n(\boldsymbol{u}))$ in the left and right images, respectively. The probabilities are computed for every combination of pixel pairs over a range of vertical and horizontal disparities. This range is chosen heuristically.

Incorporating both the optimal tessellation and the concept of epipolar spaces into this algorithm can increase its accuracy and decrease its computational complexity. First, the concept of epipolar spaces allows the disparities over which the probabilities are assessed to be chosen exactly, instead of heuristically. This ensures maximal efficiency and accuracy: probabilities are assigned to all viable combinations of pixels; combinations that are not viable are disregarded. Furthermore, using an optimally sampled image guarantees that the average number of combinations that must investigated is minimized. This reduces computational complexity.

4.6 Biological Vision

We now consider the application of optimal sampling to the human visual system (HVS). This is accomplished by determining the retinal epipolar spaces and then finding the sampling strategy that minimizes their average area. This sampling strategy can then be compared to the log-polar foveation scheme inherent in the HVS.

The following subsections formulate the retinal epipolar spaces from two distinct perspectives. The first subsection uses information gathered through psychophysical experiments. The second examines the epipolar spaces produced by the observation of natural scenes.

4.6.1 Psychophysical Construction of Epipolar Spaces

Given a point on the retina in one eye, an epipolar space represents the spatial region in the other eye that we will search for the matching point. The HVS implicitly defines these regions by establishing ranges of horizontal and vertical disparities over which signals can be corresponded. In this section we use the results of psychophysical experiments to help establish these disparity bounds.

The disparities over which humans can ascertain depth information is called the region of *qualitative stereopsis*. This is often confused with Panum's fusion area which denotes the range of disparities over which stereo images can be fused (i.e. not produce double images). For a comprehensive review of these topics see [58]. The disparity limits that define the regions of qualitative stereopsis are those most analogous to epipolar spaces. Consequently, we will examine the research pertaining to the upper disparity limits of qualitative stereopsis.

Blakemore [14] provides the most quantitative information about the horizontal disparity limits of retinal epipolar spaces. Unfortunately, his measurements are extremely sparse. These measurements are shown in Table 4.2. This table illustrates the divergent and convergent horizontal disparity limits for three different peripheral angles. Divergent and convergent disparities represent points in space lying farther and closer than the horopter, respectively. Obviously, this table does not provide the desired comprehensive functional relationship between retinal position and horizontal disparity limits. For the purposes of this research we will assume that all additional information necessary to generate the horizontal bounds of the epipolar spaces can be linearly interpolated or extrapolated from the three nearly collinear sample points in Table 4.2.

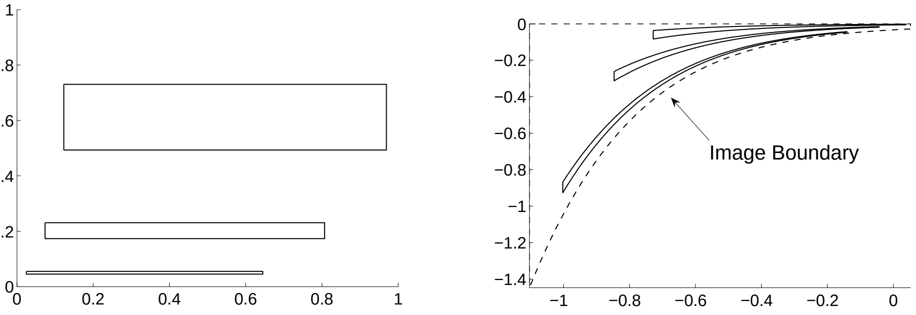
	Peripheral Angle		
	0 deg	5 deg	10 deg
Convergent Disparity	7 deg	9 deg	13 deg
Divergent Disparity	9 deg	12 deg	> 14 deg

Table 4.2: Horizontal disparity limits at different peripheral angles [14].

Though a reasonable amount of research pertains to the identification of vertical disparities limits [33, 34, 113, 117], they do not directly address the determination of the boundaries of qualitative stereopsis. Consequently, the information available is even less useful than that for horizontal disparities. The only definite conclusion that can be drawn from the literature is that matching points are not restricted to epipolar lines [113]. This indicates that the vertical limits are greater than zero. Unfortunately, we require a function relating vertical disparity limits to retinal coordinates, which the literature simply does not provide. To obtain the necessary vertical disparities limits we will assume that they are identical to those resulting from ocular rotation angles that span ±15 degrees (from straight ahead). This is a reasonable range for human stereopsis. Refer to Section 4.2 for the connection between ocular angles and vertical disparity limits.

Having established the vertical and horizontal disparity ranges, we can now formulate the epipolar spaces for a retina approximated by a flat image plane. These spaces are shown in Fig. 4.9(a). Their optimal transformation is well approximated by the following:

$$\boldsymbol{\gamma}\left(u,v\right)=\left[\beta_u\ln\left(u+\alpha\right),\,\frac{\beta_v\ln v}{\ln c\left(u\right)}\right]^t, \tag{4.50}$$



(a) (b)

Figure 4.9: Epipolar spaces for human retina before and after optimal transformation.

where β_u and β_v are constants chosen to preserve total area and $\alpha = 0.11$. The transformation of the epipolar spaces in Fig. 4.9(a) are illustrated in Fig. 4.9(b). The optimal tessellation for (4.50) is shown in Fig. 4.10(a). This can be visually compared to the log-polar sampling scheme inherent in the HVS as shown in Fig. 4.10(b). The optimal transformation warps the retinal epipolar spaces such that their average area is nearly 17 times smaller than that of uniform sampling. For log-polar foveation the performance gain over uniform sampling is 2.5.

The optimal tessellation seems, at best, a crude approximation of the log-polar polar foveation strategy "chosen" by the HVS. The question remains as to whether this discrepancy is a result of different optimization criteria or instead a consequence of the coarse approximation of the retinal epipolar spaces. Unarguably, the approximation is poor. The lack of comprehensive psychophysical models is the most significant factor. Additionally, assump-

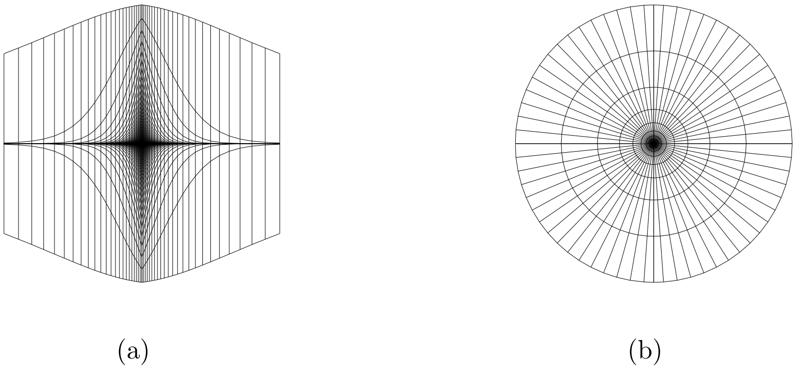


Figure 4.10: Optimal tessellation and log-polar foveation of HVS.

tions such as a flat image plane and the inevitable disregard of secondary disparity related factors such a cyclotorsional noise [43] contribute to inaccuracies in the approximation. In general, the cause of the differences between the optimal tessellation and log-polar foveation remains unanswered. Perhaps this research will provide a new impetus and paradigm for performing experiments that will adequately quantify the disparity range of qualitative stereopsis.

4.6.2 Construction of Epipolar Spaces from Natural Scenes

A prevalent school of thought suggests that biological evolution may be inexorably converging to the optimal stereo algorithm. Presupposing such a hypothesis we must wonder what are the metrics that are being optimized. In this chapter we demonstrated the advantages of minimizing the metric of average epipolar area. We now consider the epipolar spaces resulting from the observations of natural scenes by the human visual system and derive the transformation that minimizes them.

The formulation of epipolar spaces is contingent upon the geometry of the system begin modeled. In the case of the HVS this modeling can be greatly simplified by assuming that the ellipsoidal retina of the HVS can be adequately approximated by a flat image plane. With this assumption the stereo geometry can be described by the four-degrees-of-freedom camera configuration shown in Fig. 4.1; and the derivation of the epipolar spaces directly follows from the formulation presented in Section 4.2. This formulation requires a predefined range of rotation angles θ_l and θ_r for each eye. The typical range of these values for the HVS is $\theta_l, \theta_r \in [5\pi/12, 7\pi/12]$. That is, the eyes are constrained to be within ±15 degrees of straight ahead. This does not mean that the eyes can not physically rotate farther, only that if they do the visual system will not be able to extract stereo information from the retinal images.

In the derivation presented in Section 4.2 the possible horizontal disparities were confined to the interval $[-D, D]$ as delineated in (4.8). This interval remained constant irrespective of the location on the image plane. Experiments by Cormack, Liu, and Bovik [32] demonstrate that the horizontal retinal disparities produced from the observation of natural scenes are better modeled by intervals that increase in length with increasing horizontal distance u from the image center. The solid lines in Fig. 4.11 illustrate the extent of these intervals as a function of horizontal distance. These bounds were computed using the 15^{th} and 85^{th} percentile of all horizontal disparity measurements sampled at each particular u. That is, for given a point in one

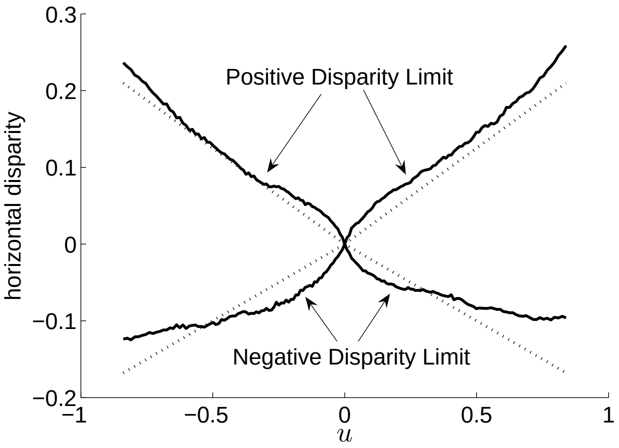


Figure 4.11: The 15^{th} and 85^{th} percentile for horizontal disparity estimates as a function of the horizontal position u.

image the horizontal separation between its retinal location and the retinal location of its matching point in the other eye will (70% of the time) fall within the bounds delineated by the curves in Fig. 4.11. The value of 70% was chosen to balance the desire to be able to extract stereo information from as many points in a scene as possible against the need to restrict the range of disparities that must be searched (thus reducing the size of the brain needed to search them in a reasonable time). The dotted lines in Fig. 4.11 provide a linear fit to the data. The importance of this linear approximation will be discussed subsequently. Having established the necessary conditions, we can now illustrate the attendant epipolar spaces. These are shown in Fig. 4.12. The solid lines represent the epipolar spaces corresponding to the exact curves presented in Fig. 4.11. The epipolar spaces dilenated with dotted lines are associated with the linear approximations from Fig. 4.11.

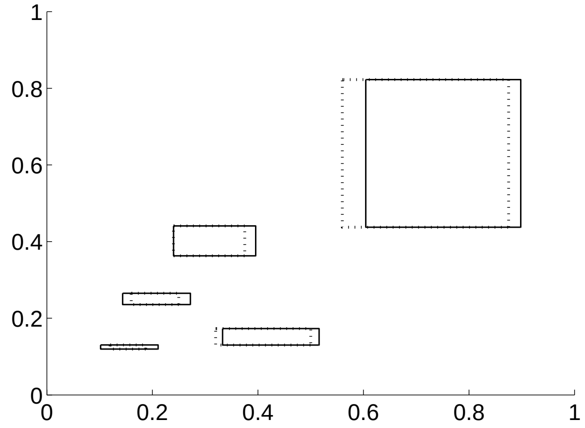


Figure 4.12: Epipolar spaces. The solid lines represent the epipolar spaces produced using the exact curves in Fig. 4.11. The dashed lines assume the linear fit.

The next step is to ascertain the mapping that minimizes the average area per epipolar space. The theorems in Section 4.3 indicate that if a transformation can be found that maps the nonuniform epipolar spaces into regions of uniform area it will be optimal with respect to their average area. Unfortunately, these theorems do not guarantee that such a mapping exists nor do they indicate how to find it. Consequently, determining an optimal transformation for the epipolar spaces resulting from the curves illustrated in Fig. 4.11 is untenable. However, we can find the optimal mapping using the linear approximation. This warping function is given by

$$\boldsymbol{\gamma}(u,v) = \left[\beta_u \ln u, \frac{\beta_v \ln v}{\ln c(u)}\right]^t, \tag{4.51}$$

where β_u and β_v are constants chosen to preserve total area. The definition of $c(u)$ was given by (4.13). Figure 4.13 illustrates the result of transforming the spaces in Fig. 4.12 using (4.51). Compared to uniform sampling the optimal sampling scheme reduces the average epipolar space by a factor of 133. Since the linear fit is only approximate, the reduction factor for the precise epipolar spaces will not achieve this value. The actual decrease in average epipolar area is 28 times.

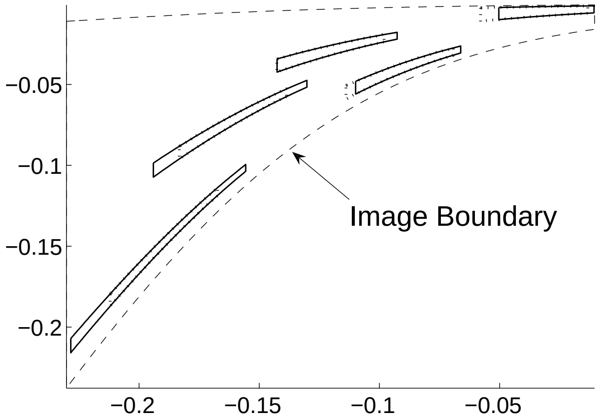


Figure 4.13: Epipolar spaces in Fig. 4.12 after transformation.

Figure 4.14(a) shows the optimal tessellation. Compare this to the log-polar sampling scheme used by the HVS [105] that is shown in Fig. 4.14(b). Similarly, both schemes increase the size of a pixel as the distance from the center increases. Contrastingly, the optimal transformation has singularities on both the u and v axes that are not present in the log-polar sampling. In terms of average epipolar area, the optimal sampling scheme performs 14 times better than the log-polar sampling scheme. However, the log-polar scheme out performs uniform sampling by a factor of two.

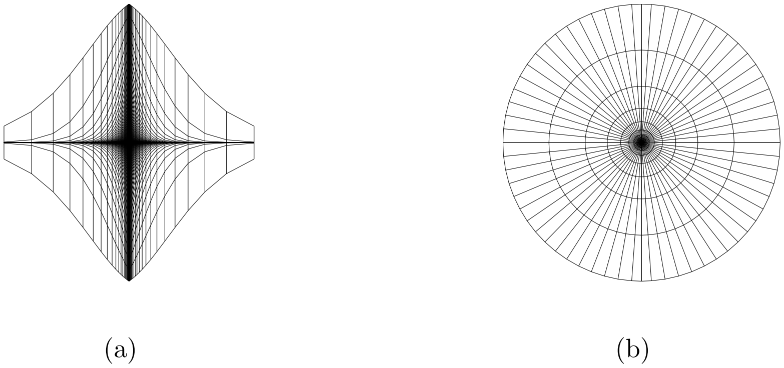

(a) (b)

Figure 4.14: Optimal tessellation and log-polar foveation.

The similarities between log-polar and the optimal transformation given by (4.51) tend to suggest that the HVS may be optimizing similar criteria. The question remains as to why the sampling schemes are different. These differences may be a result of several factors. Foremost, our assumptions in determining the size and shape of the epipolar spaces may be too restrictive. For example, the modeling of the retina as a flat image plane instead of a ellipsoid could significantly alter the optimal tessellation. Additionally, the HVS likely evolved toward a solution that was advantageous with respect to many criteria, not just average epipolar area. Perhaps extending the optimization presented in this dissertation to include other properties advantageous to biological visual systems may further shed light on this subject. In any event, a more extensive examination of the connections between optimal sampling and biological vision is definitely warranted.

4.7 Conclusion

In this chapter we introduced the concept of an epipolar space. For a point in one image, the epipolar space was defined as the region in the other image formed from the union of all associated epipolar lines produced over all possible geometric configurations. Epipolar spaces eliminate the need for calibration, but at the expense of a greater search space. To reduce this search space we introduced a nonuniform sampling scheme that was shown to be optimal with respect to the mean area of an epipolar space. Furthermore, we indicated how this information could enhance the performance of an active binocular vision system that employs an uncalibrated registration algorithm suitable for foveated images. Finally, we explored possible connections between optimal sampling and the human visual system.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Active binocular vision systems provide a powerful means for recovering depth from a complex scene. Two essential and interrelated components in this recovery process are image registration and sampling. In this dissertation we provided mathematically rigorous expatiation of both, advancing their understanding and effectiveness.

5.1.1 Registration

Image registration is the process of matching points between images. Biologically motivated phase-based techniques have been shown to provide an effective means for correspondence. Unfortunately, they are difficult to analyze mathematically, and consequently, have resulted in a variety of heuristic implementations. In this dissertation we were able to obviate these heuristics by introducing two new paradigms. These paradigms increased the understanding and performance of phase-based registration algorithms by providing firm mathematical frameworks.

First, we considered the first and second derivatives of local phase and their importance in detecting regions of nonlinear phase. We began by deriving probability density functions for these derivatives under the assumption of Gaussian white noise. We then considered the effects of constraining their domains to regions of linear phase. In order to identify these regions of linear phase, we proposed using the second derivative of phase. We then described how prevalent methods proposed in the literature can be seen as approximations of the second derivative. We empirically validated these assertions and their extensibility by comparing the different methods on white noise, a natural scene, and a random-dot stereo pair.

We next addressed the difficulty of multiscale fusion in phase differencing by introducing a novel registration technique that incorporated strategies of both phase differencing and local correlation. This hybrid approach allowed us to combine the rapidly calculated, high resolution results associated with phase differencing with the multiscale information available to local correlation. Finally, we demonstrated the performance on two pairs of stereo images.

5.1.2 Foveation

Foveation is an essential component to depth recovery for active binocular vision systems. Experiments in biological vision have encouraged the use of space-variant sampling schemes like log-polar. Most space-variant sampling schemes used for binocular vision tasks were developed elsewhere and then applied to areas of stereo vision. In this dissertation we demonstrated how foveation *results* from allocating samples in an optimal fashion.

First, we introduced the concept of an epipolar space. For a point in one image, the corresponding epipolar space was defined as the region in the other image formed from the union of all associated epipolar lines produced over all possible geometric configurations. Epipolar spaces eliminate the need for calibration, but at the expense of a greater search space. Since it is necessary to consider the entire epipolar space when searching for a matching point, the computational complexity and accuracy of the registration algorithm is directly related to the sizes of the epipolar spaces.

In order to minimize computational complexity and maximize accuracy, we introduced a nonuniform sampling scheme that was shown to be optimal with respect to the mean area of an epipolar space. This derivation was accomplished by the the development of several proofs that demonstrated that when nonuniformly sized regions are transformed into regions of uniform area, their average area is minimized. From the optimal sampling scheme we constructed a tessellation that could be used to enhance the performance of any active binocular vision system that employs an uncalibrated registration algorithm suitable for foveated images.

5.2 Future Work

Though both registration and foveation are separate components in the depth recovery process, their relation is highly interdependent. Nearly all registration algorithms presently available are only compatible with uniformly sampled images. Registering foveated images requires developing specifically tailored correspondence techniques. Which current methodologies might best be adapted for use with foveated images is an open question.

In this section we address the difficultly of registering nonuniformly sampled images, demonstrating why most current registration techniques are inherently incompatible with foveated images. Next we discuss how phase differencing methods like those presented in this dissertation may provide the necessary solution. Finally, we indicate steps needed to begin the implementation of such a solution.

5.2.1 Registration and Nonuniform Sampling

The registration of nonuniformly sampled images is more complex than the registration of uniformly sampled images. This becomes obvious if we consider the simple hypothesis that stereo images are related by local shifts. With vergent geometries these shifts can be both vertical and horizontal. For uniformly sampled images the shifts are relatively innocuous, manifesting as local displacements. Consequently, any patch in one image can be matched to a near identical patch in the other. Foveation obfuscates matters. For a patch in one image the corresponding patch in the other is represented by a *different number of pixels* whose values have been *blurred to differing degrees.*

Figure 5.1 highlights the difficulty in matching image patches in nonuniformly sampled images such as those created by the log-polar transformation. In Figure 5.1 two square image patches are shown in the square grid on the right. A shift in this cortical coordinate system produces a rotation in the retinal coordinates. If were were to perform a shift in the retinal coordinates (which would be a typical occurrence in stereo pairs) the corresponding regions in the cortical image would be oddly shaped regions of different sizes. Consequently, comparing the two regions would be very difficult.

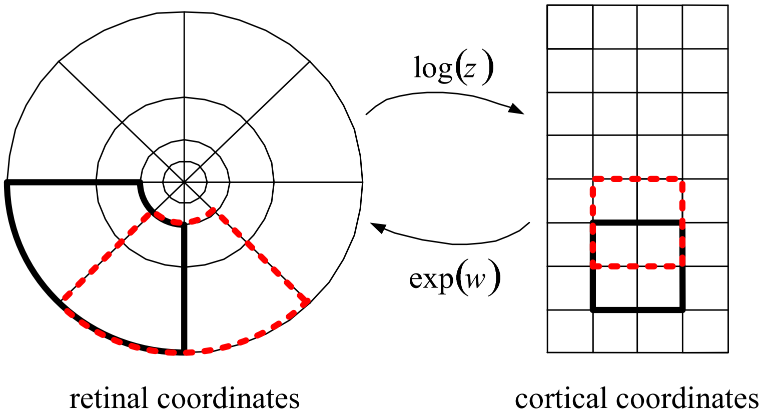


Figure 5.1: Log-polar mapping. This figure demonstrates that a shift in cortical coordinates is not equivalent to shift in retinal coordinates.

These factors prove too difficult to overcome for many types of registration techniques, such as correlation based algorithms that attempt to directly match image patches. Perhaps this explains the dearth of dense stereo registration algorithms for foveated images. In fact, only two algorithms work with sampling schemes that have tessellations that are not composed of squares [12, 102]. The first of these requires knowledge of the exact stereo geometry, making it inappropriate for an uncalibrated ABVS.

5.2.2 Phase-Based Registration and Nonuniform Sampling

Local phase may provide a solution for the difficulties inherent in the registration of nonuniformly sampled images. Since local phase in known to be quasi-linear with respect to the local disparity, depth can be recovered in a single step from the phase difference between *spatially identical* regions in each image. Since the phase measurements are made using pixels at the same locations, the problem of comparing patches with different numbers of pixels is eliminated. The problem of unequal blur persists. This is due to the fact that the foveation occurs after the local shifts. Anecdotally, there is evidence to suggest that phase is robust to such differences in blur. In [23], Chen foveated a pair of stereo images about the same arbitrary point and then measured the global phase for both stereo images. Using this phase information he was able to determine the local disparity at the point of foveation. Repeating this process at every point, he constructed a dense disparity map of relatively high accuracy. Additionally, a good deal a research has shown success using foveated images for vergence [12,20,74]. Since the entire image, and not image patches, is considered when computing similarity measures, the identical set of pixels is used. These correlation based approaches seem to have little problem with the unequal blur. In any case, if blurring remains a persistent problem, additional local filtering could be applied to ensure that all points are blurred equally.

On a final note biological ABVS are known to sample the environment nonuniformly, i.e. they have foveated retinas. Interestingly, there is also strong evidence suggesting that they use local phase information [65]. Phase-differencing may even be a technique employed by the human brain [38].

5.2.3 The Next Steps

The next steps in the evolution of this research would be to construct a phase-based registration algorithm capable of being used with the optimal nonuniform sampling scheme developed in this dissertation. Though substantial groundwork has been established for this task, a considerable amount of work remains:

1. The quadrature filters used to extract local phase would need to be adapted for application in the warped domain (the warped domain is analogous to the log-polar cortical domain). Works by Clark [28] and Zeevi [134] provide a basis for this extension. These works primarily concern antialiasing. Since the purpose of a registration algorithm is not to reconstruct the images from their samples, but to correspond the images *without time-consuming reconstruction*, adaptations are necessary for appropriate signal preservation.

2. The spectral ranges and spatial scales of filters used at each location must be varied depending upon the local resolution. For example, in the periphery where the sampling is coarse, high frequency filters will no longer be appropriate.

3. In this dissertation we stipulated a constant maximum horizontal dis-

parity. This constraint restricted the depth about the horopter in which objects could be matched. It is quite possible that this value should not be constant, but should increase with the distance from the center of the image. Since it is assumed that depth recovery occurs after the cameras are in vergence, the center points of both images contain projections of the exact same 3D point, and consequently, have zero disparity. As we move farther away from the center we can expect the objects to have greater variations in depth, increasing the expected horizontal disparities. Modifying the allowable horizontal disparities alters the sizes of the epipolar spaces. The results of this dissertation provide a clear means for deriving an optimal sampling scheme for these newly formulated epipolar spaces.

4. Test data must be acquired for the specific range of geometries assumed when constructing the optimal sampling scheme.

5. The optimal sampling scheme is not unique; there is a family optimal sampling strategies. Future research could use other constraints to select the member of this family that optimizes some additional criterion. Such a criterion might be 3D reconstruction error [9, 42].

6. Since foveated images have large pixels in the periphery, the effect of filtering across image boundaries becomes more pronounced, i.e. it can disrupt large portions of the image. Any phase-based technique would need an a reliable technique for recovering phase information near the

edges of the image.

Appendix

Appendix 1

Local Phase Derivations

1.1 Probability Density Function Derivations

Consider the zero mean Gaussian random variables $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$, $\boldsymbol{e}$, and $\boldsymbol{f}$. Let K defined in (3.21) be the covariance matrix relating $\boldsymbol{a}$, $\boldsymbol{c}$, and $\boldsymbol{e}$ and also relating $\boldsymbol{b}$, $\boldsymbol{d}$, and $\boldsymbol{f}$. These matrices represent the only dependencies between the six random variables (i.e. $\boldsymbol{a}$, $\boldsymbol{c}$, and $\boldsymbol{e}$ are independent with respect to $\boldsymbol{b}$, $\boldsymbol{d}$, and $\boldsymbol{f}$). Their joint probability density function (pdf) is

$$f(a,b,c,d,e,f) = \left[(2\pi)^3 \left|K\right|\right]^{-1} e^{-\frac{1}{2}[a\ d\ e]K^{-1}[a\ d\ e]^T} e^{-\frac{1}{2}[b\ c\ f]K^{-1}[b\ c\ f]^T}, \qquad (1.1)$$

where $|K|$ is the determinant of K. In the following subsections we will consider the properties of functional combinations of these random variables.

1.1.1 Probability Density Functions for $\boldsymbol{\xi}$ and $\boldsymbol{\tau}$

The goal of this section is to find the joint and marginal pdfs of the random variables $\boldsymbol{\xi}$ and $\boldsymbol{\tau}$, where $\boldsymbol{\xi} = \frac{\boldsymbol{ad}-\boldsymbol{bc}}{\boldsymbol{a}^2+\boldsymbol{b}^2}$ and $\boldsymbol{\tau} = \frac{\boldsymbol{af}-\boldsymbol{be}}{\boldsymbol{a}^2+\boldsymbol{b}^2}$. Consider the following functions of the random variables $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$, $\boldsymbol{e}$, and $\boldsymbol{f}$: $\boldsymbol{w} = \boldsymbol{ad} - \boldsymbol{bc}$, $\boldsymbol{x} = \boldsymbol{a}^2 + \boldsymbol{b}^2$, $\boldsymbol{y} = \boldsymbol{a}^2$, $\boldsymbol{z} = \boldsymbol{bc}$, $\boldsymbol{r} = \boldsymbol{af} - \boldsymbol{be}$, and $\boldsymbol{s} = \boldsymbol{be}$. The solutions of this system of equations are $a = \pm\sqrt{y}$, $b = \pm\sqrt{x-y}$, $c = \frac{x}{\pm\sqrt{x-y}}$, $d = \frac{w+z}{\pm\sqrt{y}}$, $e = \frac{s}{\pm\sqrt{x-y}}$, and $f = \frac{r+s}{\pm\sqrt{y}}$. Because of the $\pm$ nature of the solutions $\pm\sqrt{y}$ and

$\pm\sqrt{x-y}$, there are four unique roots (combination of solutions). The Jacobian of the these transformations is $J(w,x,y,z,r,s) = 4\left[y\left(x-y\right)\right]^{\frac{3}{2}}$. Incorporating these results into (1.1) and simplifying, the joint pdf can be expressed as follows (for simplicity let $K^{-1} = \begin{bmatrix} 2\alpha & 0 & \gamma \\ 0 & 2\beta & 0 \\ \gamma & 0 & 2\delta \end{bmatrix}$):

$$\begin{aligned}
f(w,x,y,z,r,s) &= \frac{\left[y\left(x-y\right)\right]^{-\frac{3}{2}}}{\left(2\pi\right)^3 |K|} e^{-\frac{1}{2}[\sqrt{y} \quad \frac{z}{\sqrt{x-y}} \quad \frac{s}{\sqrt{x-y}}]K^{-1}[\sqrt{y} \quad \frac{z}{\sqrt{x-y}} \quad \frac{s}{\sqrt{x-y}}]^t} \times \\
&\qquad e^{-\frac{1}{2}[\sqrt{x-y} \quad \frac{w+z}{\sqrt{y}} \quad \frac{r+s}{\sqrt{y}}]K^{-1}[\sqrt{x-y} \quad \frac{w+z}{\sqrt{y}} \quad \frac{r+s}{\sqrt{y}}]^t} \\
&= \frac{\left[y\left(x-y\right)\right]^{-\frac{3}{2}}}{\left(2\pi\right)^3 |K|} e^{-\left[z^2\left(\frac{\beta}{x-y}+\frac{\beta}{y}\right)+z\left(\frac{2\beta w}{y}\right)\right]} \times \\
&\qquad e^{-\left[\alpha x+\beta\frac{w^2}{y}+\delta\left(\frac{s^2}{x-y}+\frac{(r+s)^2}{y}\right)+\gamma\left(\frac{s\sqrt{y}}{\sqrt{x-y}}+\frac{(r+s)\sqrt{x-y}}{\sqrt{y}}\right)\right]} \\
&= \frac{\left[y\left(x-y\right)\right]^{-\frac{3}{2}}}{\left(2\pi\right)^3 |K|} e^{-\alpha x} e^{-\frac{\beta[z+\mu_z]^2}{(x-y)y}} \times \\
&\qquad e^{-\frac{1}{(x-y)y}\left[-\beta\frac{w^2(x-y)^2}{x}+\beta w^2(x-y)\right]} \times \\
&\qquad e^{-\frac{1}{(x-y)y}\left[\delta\left(ys^2+[x-y][r+s]^2\right)+\gamma\sqrt{y(x-y)}(ys+[x-y][r+s])\right]} \\
&= \frac{\left[y\left(x-y\right)\right]^{-\frac{3}{2}}}{\left(2\pi\right)^3 |K|} e^{-\alpha x} e^{-\frac{\beta[z+\mu_z]^2}{(x-y)y}} \times \\
&\qquad e^{-\frac{\delta x}{(x-y)y}\left[s^2+s\left(\frac{2r[x-y]}{x}+\frac{\gamma\sqrt{y[x-y]}}{\delta}\right)\right]} e^{-\frac{1}{y}\left[\beta\frac{w^2 y}{x}+\delta r^2+\gamma r\sqrt{y(x-y)}\right]} \\
&= \frac{\left[y\left(x-y\right)\right]^{-\frac{3}{2}}}{\left(2\pi\right)^3 |K|} e^{-\alpha x} e^{-\frac{\beta[z+\mu_z]^2}{(x-y)y}} e^{-\frac{\delta x[s+\mu_s]^2}{(x-y)y}} \times \\
&\qquad e^{-\frac{1}{4\delta x}\left[-\gamma^2x^2+4\delta^2r^2+4\beta\delta w^2\right]} \\
&= \frac{\left[y\left(x-y\right)\right]^{-\frac{3}{2}}}{\left(2\pi\right)^3 |K|} e^{-\alpha x-\frac{\beta[z+\mu_z]^2}{(x-y)y}-\frac{\delta x[s+\mu_s]^2}{(x-y)y}-\frac{-\gamma^2x^2+4\delta^2r^2+4\beta\delta w^2}{4\delta x}}
\end{aligned}$$

where $\mu_z = \frac{w(x-y)}{x}$ and $\mu_s = \frac{2\delta(x-y)r+\gamma x\sqrt{y(x-y)}}{2\delta x}$. We next integrate out the variables z, s, and y:

$$\begin{aligned}
f(w,x,y,r) &= \frac{[y\,(x-y)]^{-\frac{3}{2}}}{(2\pi)^3\,|K|} e^{-\alpha x} e^{-\frac{1}{4\delta x}\left[-\gamma^2x^2+4\delta^2r^2+4\beta\delta w^2\right]} \times \\
&\qquad \int_{-\infty}^{\infty} e^{-\frac{\beta[z+\mu_z]^2}{(x-y)y}}\,dz \int_{-\infty}^{\infty} e^{-\frac{\delta x[s+\mu_s]^2}{(x-y)y}}\,ds \\
&= \frac{[y\,(x-y)]^{-\frac{1}{2}}}{8\pi^2|K|x\sqrt{\beta\delta}} e^{-\alpha x} e^{-\frac{1}{4\delta x}\left[-\gamma^2x^2+4\delta^2r^2+4\beta\delta w^2\right]} \\
f(w,x,r) &= \frac{1}{8\pi^2|K|x\sqrt{\beta\delta}} e^{-\alpha x} e^{-\frac{1}{4\delta x}\left[-\gamma^2x^2+4\delta^2r^2+4\beta\delta w^2\right]} \int_0^x \frac{dy}{\sqrt{y(x-y)}} \\
&= \frac{1}{8\pi|K|x\sqrt{\beta\delta}} e^{-\alpha x} e^{-\frac{1}{4\delta x}\left[-\gamma^2x^2+4\delta^2r^2+4\beta\delta w^2\right]}.
\end{aligned}$$

Continuing, consider three new functions of the random variables $\boldsymbol{x}$, $\boldsymbol{w}$, and $\boldsymbol{r}$: $\boldsymbol{\xi} = \boldsymbol{w}/\boldsymbol{x}$, $\boldsymbol{\lambda} = \boldsymbol{x}$, and $\boldsymbol{\tau} = \boldsymbol{r}/\boldsymbol{x}$. (Since the mathematics may obfuscate matters we reiterate that these representations of $\boldsymbol{\xi}$ and $\boldsymbol{\tau}$ correspond to the original definitions: $\boldsymbol{\xi} = \frac{\boldsymbol{ad}-\boldsymbol{bc}}{\boldsymbol{a}^2+\boldsymbol{b}^2}$ and $\boldsymbol{\tau} = \frac{\boldsymbol{af}-\boldsymbol{be}}{\boldsymbol{a}^2+\boldsymbol{b}^2}$). The Jacobian is $J(\xi,\lambda,\tau) = 1/\lambda^2$. Finally, we can substitute the solutions $w = \xi\lambda$, $x = \lambda$, and $r = \tau\lambda$ into the previous equation for $f(w,x,r)$ and integrate with respect to λ to arrive at the joint pdf of $\boldsymbol{\xi}$ and $\boldsymbol{\tau}$.

$$\begin{aligned}
f(\xi,\tau) &= \frac{1}{8\pi|K|\sqrt{\beta\delta}} \int_0^{\infty} \lambda e^{-\frac{\lambda}{4\delta}\left[4\delta\alpha-\gamma^2+4\delta^2\tau^2+4\beta\delta\xi^2\right]} d\lambda \\
&= \frac{\delta^{\frac{3}{2}}}{8\pi|K|\sqrt{\beta}} \frac{1}{\left(-\frac{\gamma^2}{4}+\delta\alpha+\delta^2\tau^2+\beta\delta\xi^2\right)^2} \\
&= \frac{|K|^{\frac{3}{2}}}{\pi\sigma_a^5\sigma_c^4}\left[\frac{|K|}{\sigma_a^4\sigma_c^2}\left(\frac{\sigma_c^2}{\sigma_a^2}+\xi^2\right)+\tau^2\right]^{-2} \qquad (1.2)
\end{aligned}$$

Integrating (1.2) with respect to τ and ξ produces the distributions $f\,(\xi)$ and

$f\left(\tau\right)$, respectively:

$$f\left(\xi\right)=\frac{\frac{\sigma_c^2}{\sigma_a^2}}{2\left(\frac{\sigma_c^2}{\sigma_a^2}+\xi^2\right)^{\frac{3}{2}}} \tag{1.3}$$

$$f\left(\tau\right)=\frac{\frac{|K|}{\sigma_a^4\sigma_c^2}}{2\left(\frac{|K|}{\sigma_a^4\sigma_c^2}+\tau^2\right)^{\frac{3}{2}}}. \tag{1.4}$$

Some important moments of $\boldsymbol{\xi}$ and $\boldsymbol{\tau}$ are $E[\boldsymbol{\xi}]=0$, $E[\boldsymbol{\tau}]=0$, $E[\boldsymbol{\xi}^2]\to\infty$, and $E[\boldsymbol{\tau}^2]\to\infty$. Noting that $\int_{-\infty}^{\infty}|t|\frac{t_o}{2(t_o+t^2)^{3/2}}dt=\sqrt{t_o}$, we see

$$E[|\boldsymbol{\xi}|]=\sqrt{\frac{\sigma_c^2}{\sigma_a^2}} \tag{1.5}$$

$$E[|\boldsymbol{\tau}|]=\sqrt{\frac{|K|}{\sigma_a^4\sigma_c^2}}. \tag{1.6}$$

1.1.2 Alternative Representation for $\boldsymbol{\tau}$

In the representation of $\boldsymbol{\tau}$ shown in (3.10) the random variables $\boldsymbol{a}$ and $\boldsymbol{e}$ are dependent, as are $\boldsymbol{b}$ and $\boldsymbol{f}$. These dependencies make further derivations involving $\boldsymbol{\tau}$ more difficult. In this section we introduce an alternative form in which these dependencies are not present. Begin by letting $\boldsymbol{s}=\alpha_1\boldsymbol{a}+\boldsymbol{e}$ and $\boldsymbol{t}=\alpha_2\boldsymbol{b}+\boldsymbol{f}$. We desire values for α_1 and α_2 such that $\boldsymbol{a}$ is independent of $\boldsymbol{s}$ and $\boldsymbol{b}$ is independent of $\boldsymbol{t}$. This requires $\mathrm{E}[\boldsymbol{as}]=\mathrm{E}[\boldsymbol{a}\left(\alpha_1\boldsymbol{a}+\boldsymbol{e}\right)]=\alpha_1\mathrm{E}[\boldsymbol{a}^2]+\mathrm{E}[\boldsymbol{ae}]=0$ and $\mathrm{E}[\boldsymbol{bt}]=\alpha_2\mathrm{E}\left[\boldsymbol{b}^2\right]+\mathrm{E}[\boldsymbol{bf}]=0$, i.e.

$$\alpha_1=-\frac{\mathrm{E}[\boldsymbol{ae}]}{\boldsymbol{a}^2}=\frac{\sigma_c^2}{\sigma_a^2}$$

$$\alpha_2=-\frac{\mathrm{E}[\boldsymbol{bf}]}{\boldsymbol{b}^2}=\frac{\sigma_c^2}{\sigma_a^2}.$$

Using these values for α_1 and α_2, the variances of $\mathbf{s}$ and $\boldsymbol{t}$ become

$$\sigma_s^2 = \mathrm{E}\left[\left(\frac{\sigma_c^2}{\sigma_a^2}\boldsymbol{s} + \boldsymbol{e}\right)^2\right] = \frac{|K|}{\sigma_a^4\sigma_c^2} = \sigma_t^2. \tag{1.7}$$

Furthermore, the random variables $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, $\boldsymbol{d}$, $\boldsymbol{s}$, and $\boldsymbol{t}$ are all uncorrelated and, consequently, independent. (Refer to Section 3.3.1 if this is not clear.)

Continuing we can now express $\boldsymbol{\tau}$ as follows:

$$\boldsymbol{\tau} = \frac{\boldsymbol{af} - \boldsymbol{be}}{\boldsymbol{a}^2 + \boldsymbol{b}^2} = \frac{\boldsymbol{a}\left(\alpha_2\boldsymbol{b} + \boldsymbol{t}\right) - \boldsymbol{b}\left(\alpha_1\boldsymbol{a} + \boldsymbol{s}\right)}{\boldsymbol{a}^2 + \boldsymbol{b}^2} = \frac{\boldsymbol{at} - \boldsymbol{bs}}{\boldsymbol{a}^2 + \boldsymbol{b}^2} \tag{1.8}$$

Because of the independence of $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{s}$, and $\boldsymbol{t}$, this form of $\boldsymbol{\tau}$ behaves similarly to $\boldsymbol{\xi}$ (where $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ are independent) with $\boldsymbol{s}$ and $\boldsymbol{t}$ replacing $\boldsymbol{c}$ and $\boldsymbol{d}$. Consequently, equations (1.3) and (1.5) hold valid for $\boldsymbol{\tau}$ after replacing σ_c with the σ_s in (1.7). This substitution reproduces (1.4) and (1.6) as it should.

1.2 Polar Forms for $\boldsymbol{\xi}$, $\boldsymbol{\tau}$, $\boldsymbol{\chi}$, and $\boldsymbol{\nu}$

In this section we derive the polar forms of $\boldsymbol{\xi}$, $\boldsymbol{\tau}$, $\boldsymbol{\chi}$, and $\boldsymbol{\nu}$ and demonstrate their use in determining the joint pdf relating $\boldsymbol{\nu}$ and $\boldsymbol{\xi}$. Consider the constituent components of (3.10): $\boldsymbol{\tau} = \frac{\boldsymbol{af}-\boldsymbol{be}}{\boldsymbol{a}^2+\boldsymbol{b}^2} = \frac{\boldsymbol{at}-\boldsymbol{bs}}{\boldsymbol{a}^2+\boldsymbol{b}^2}$ (see Appendix 1.1.2) and $\boldsymbol{\nu} = \boldsymbol{\xi}\boldsymbol{\chi} = \frac{\boldsymbol{ad}-\boldsymbol{dc}}{\boldsymbol{a}^2+\boldsymbol{b}^2}\frac{\boldsymbol{ac}+\boldsymbol{bd}}{\boldsymbol{a}^2+\boldsymbol{b}^2}$, where $\boldsymbol{\xi} = \frac{\boldsymbol{ad}-\boldsymbol{dc}}{\boldsymbol{a}^2+\boldsymbol{b}^2}$ and $\boldsymbol{\chi} = \frac{\boldsymbol{ac}+\boldsymbol{bd}}{\boldsymbol{a}^2+\boldsymbol{b}^2}$. We begin by converting each into polar coordinates, setting $\boldsymbol{a} = \boldsymbol{r_1}\cos\left(\boldsymbol{\phi_1}\right)$, $\boldsymbol{b} = \boldsymbol{r_1}\sin\left(\boldsymbol{\phi_1}\right)$, $\boldsymbol{c} = \boldsymbol{r_2}\cos\left(\boldsymbol{\phi_2}\right)$, $\boldsymbol{d} = \boldsymbol{r_2}\sin\left(\boldsymbol{\phi_2}\right)$, $\boldsymbol{c} = \boldsymbol{r_3}\cos\left(\boldsymbol{\phi_3}\right)$, and $\boldsymbol{d} = \boldsymbol{r_3}\sin\left(\boldsymbol{\phi_3}\right)$. The random variables $\boldsymbol{r_1}$, $\boldsymbol{r_2}$, and $\boldsymbol{r_3}$ have Rayleigh distributions while $\boldsymbol{\phi_1}$, $\boldsymbol{\phi_2}$ and $\boldsymbol{\phi_3}$ are distributed uniformly over the interval $[-\pi : \pi]$ [52]. All the random variables are independent of each other. In polar polar coordinates we have

$$\boldsymbol{\xi} = \frac{\boldsymbol{r_2}}{\boldsymbol{r_1}}\sin\left(\boldsymbol{\phi_2} - \boldsymbol{\phi_1}\right) = \boldsymbol{R_1}\sin\boldsymbol{\Phi_1} \tag{1.9}$$

$$\boldsymbol{\chi} = \frac{\boldsymbol{r_2}}{\boldsymbol{r_1}} \cos\left(\boldsymbol{\phi_2} - \boldsymbol{\phi_1}\right) = \boldsymbol{R_1} \cos \boldsymbol{\Phi_1} \tag{1.10}$$

$$\boldsymbol{\tau} = \frac{\boldsymbol{r_3}}{\boldsymbol{r_1}} \sin\left(\boldsymbol{\phi_3} - \boldsymbol{\phi_1}\right) = \boldsymbol{R_2} \sin \boldsymbol{\Phi_2} \tag{1.11}$$

$$\boldsymbol{\nu} = \boldsymbol{\xi}\boldsymbol{\chi} = \boldsymbol{R_1}^2 \sin \boldsymbol{\Phi_1} \cos \boldsymbol{\Phi_1}, \tag{1.12}$$

where $\boldsymbol{R_1} = \boldsymbol{r_2}/\boldsymbol{r_1}$, $\boldsymbol{R_2} = \boldsymbol{r_3}/\boldsymbol{r_1}$, $\boldsymbol{\Phi_1} = \boldsymbol{\phi_2} - \boldsymbol{\phi_1}$, and $\boldsymbol{\Phi_2} = \boldsymbol{\phi_3} - \boldsymbol{\phi_1}$. $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$ are uniformly distributed over $[-\pi, \pi]$ and are independent of $\boldsymbol{R_1}$, $\boldsymbol{R_2}$, and each other. The joint distribution relating $\boldsymbol{r_1}$, $\boldsymbol{r_2}$, and $\boldsymbol{r_3}$ is

$$f\left(r_1, r_2, r_3\right) = f\left(r_1\right) f\left(r_2\right) f\left(r_3\right) = \frac{r_1 r_2 r_3}{\sigma_a^2 \sigma_c^2 \sigma_s^2} e^{-\frac{r_1^2}{2\sigma_a^2}} e^{-\frac{r_2^2}{2\sigma_c^2}} e^{-\frac{r_3^2}{2\sigma_s^2}}. \tag{1.13}$$

To find the joint pdf relating relating $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$, begin by letting $\boldsymbol{R_1} = \boldsymbol{r_2}/\boldsymbol{r_1}$, $\boldsymbol{R_2} = \boldsymbol{r_3}/\boldsymbol{r_1}$, and $\boldsymbol{z} = \boldsymbol{r_1}$. The solutions to these equations are $r_1 = z$, $r_2 = zR_1$, and $r_3 = zR_2$. The Jacobian for the transformation is z^2. Substituting these solutions into (1.13) and integrating with respect to z we have

$$\begin{aligned} f\left(R1, R2\right) &= \frac{R_1 R_2}{\sigma_a^2 \sigma_c^2 \sigma_s^2} \int_0^\infty z^5 e^{-z^2\left(\frac{1}{2\sigma_a^2} + \frac{R_1^2}{2\sigma_c^2} + \frac{R_2^2}{2\sigma_s^2}\right)} \\ &= \frac{R_1 R_2}{\sigma_a^2 \sigma_c^2 \sigma_s^2} \left(\frac{1}{2\sigma_a^2} + \frac{R_1^2}{2\sigma_c^2} + \frac{R_2^2}{2\sigma_s^2}\right)^{-1} \\ &= 8 R_1 R_2 \frac{\sigma_s^4}{\sigma_a^2 \sigma_c^2} \left(\frac{\sigma_s^2}{\sigma_c^2}\left[\frac{\sigma_c^2}{\sigma_a^2} + R_1^2\right] + R_2^2\right)^{-1}. \end{aligned} \tag{1.14}$$

Integrating out R_2 and R_1 we arrive at the following marginal distributions:

$$f\left(R_1\right) = \frac{2\sigma_c^2/\sigma_a^2 R_1}{\left(\sigma_c^2/\sigma_a^2 + R_1^2\right)^2}. \tag{1.15}$$

$$f\left(R_2\right) = \frac{2\sigma_s^2/\sigma_a^2 R_2}{\left(\sigma_s^2/\sigma_a^2 + R_2^2\right)^2}. \tag{1.16}$$

For a moment we return our attention to $\boldsymbol{\nu}$. From (1.15) we can easily show that $\mathrm{E}\left[\boldsymbol{R}_{\mathbf{1}}^{\mathbf{2}}\right]$ is not finite. As a consequence, $\mathrm{E}[|\boldsymbol{\nu}|] = \mathrm{E}\left[{\boldsymbol{R}_{\mathbf{1}}}^2\right]\mathrm{E}[|\cos\boldsymbol{\Phi}_{\mathbf{1}}\sin\boldsymbol{\Phi}_{\mathbf{1}}|]$ is also infinite as is the variance of $\boldsymbol{\nu}$.

We can now find the joint distribution $f(\nu,\xi)$. Remembering that $\boldsymbol{\xi} = \boldsymbol{R}_{\mathbf{1}}\sin\boldsymbol{\Phi}_{\mathbf{1}}$ and $\boldsymbol{\nu} = {\boldsymbol{R}_{\mathbf{1}}}^2\cos\boldsymbol{\Phi}_{\mathbf{1}}\sin\boldsymbol{\Phi}_{\mathbf{1}}$, we can find the expression for $f(\nu,\xi)$ by substituting the solutions $\Phi_1 = \cos^{-1}(\xi/\nu)$ and $R_1 = \xi\sqrt{\nu^2+\xi^4}$ (with Jacobian $J = \xi^2\sqrt{\nu^2+\xi^4}$) into $f(R_1,\Phi_1) = f(R_1)\,f(\Phi_1)$:

$$f(\nu,\xi) = \frac{\sigma_c^2/\sigma_a^2\xi^3}{\pi\left(\xi^2\left[\sigma_c^2/\sigma_a^2+\xi^2\right]+\nu^2\right)^2}. \tag{1.17}$$

1.3 Conditional PDF Derivations

In this section we analyze the effects of the constraints posed in (3.11) and (3.13) on the random variables $\boldsymbol{\xi}$, $\boldsymbol{\tau}$, and $\boldsymbol{\nu}$. The constraint shown in (3.11) restricts the possible values of $\boldsymbol{\xi}$. In order to quantify the influence of this constraint we will find expressions for $\mathrm{E}\left[|\boldsymbol{\xi}|\,\middle|\,|\xi|<\rho_1\right]$, $\mathrm{E}\left[|\boldsymbol{\tau}|\,\middle|\,|\xi|<\rho_1\right]$, and $\mathrm{E}\left[|\boldsymbol{\nu}|\,\middle|\,|\xi|<\rho_1\right]$. First we note that

$$\begin{aligned}\mathrm{E}\left[|\cdot|\,\middle|\,|\xi|<\rho\right] &= \frac{\int_{-\rho}^{\rho}\mathrm{E}\left[|\cdot|\,\middle|\,\xi\right]f(\xi)\,d\xi}{\int_{-\rho}^{\rho}f(\xi)\,d\xi}\\ &= \frac{1}{\rho}\frac{\sigma_c^2}{\sigma_a^2}\left[\frac{\sigma_c^2}{\sigma_a^2}+\rho^2\right]^{\frac{1}{2}}\int_{-\rho}^{\rho}\mathrm{E}\left[|\cdot|\,\middle|\,\xi\right]f(\xi)\,d\xi. \end{aligned}\tag{1.18}$$

The initial step is to find $\mathrm{E}\left[|\boldsymbol{\xi}|\,\middle|\,\xi\right]$, $\mathrm{E}\left[|\boldsymbol{\tau}|\,\middle|\,\xi\right]$, and $\mathrm{E}\left[|\boldsymbol{\nu}|\,\middle|\,\xi\right]$. Obviously, we have

$$\mathrm{E}\left[|\boldsymbol{\xi}|\,\middle|\,\xi\right] = \xi. \tag{1.19}$$

To determine the other expected values we first determine the conditional distribution functions of $\boldsymbol{\xi}$ and $\boldsymbol{\nu}$ with respect to $\boldsymbol{\xi}$:

$$f\left(\tau\middle|\xi\right) = \frac{f\left(\xi,\tau\right)}{f\left(\xi\right)} = \frac{2\left(\frac{|K|}{\sigma_a^2\sigma_c^4}\left[\frac{\sigma_c^2}{\sigma_a^2}+\xi^2\right]\right)^{\frac{3}{2}}}{\pi\left(\frac{|K|}{\sigma_a^2\sigma_c^4}\left[\frac{\sigma_c^2}{\sigma_a^2}+\xi^2\right]+\tau^2\right)^2} \tag{1.20}$$

$$f\left(\nu\middle|\xi\right) = \frac{f\left(\xi,\nu\right)}{f\left(\xi\right)} = \frac{2\left(\xi^2\left[\frac{\sigma_c^2}{\sigma_a^2}+\xi^2\right]\right)^{\frac{3}{2}}}{\pi\left(\xi^2\left[\frac{\sigma_c^2}{\sigma_a^2}+\xi^2\right]+\nu^2\right)^2} \tag{1.21}$$

Using the fact that $\int_{-\infty}^{\infty}\frac{2t_o^{3/2}|t|}{\pi(t_o+t^2)^2}dt = \frac{2}{\pi}\sqrt{t_o}$ we find

$$\mathrm{E}\left[|\boldsymbol{\tau}|\,\middle|\xi\right] = \frac{2}{\pi}\left(\frac{|K|}{\sigma_a^2\sigma_c^4}\left[\frac{\sigma_c^2}{\sigma_a^2}+\xi^2\right]\right)^{\frac{1}{2}} \tag{1.22}$$

$$\mathrm{E}\left[|\boldsymbol{\nu}|\,\middle|\xi\right] = \frac{2}{\pi}\left(\xi^2\left[\frac{\sigma_c^2}{\sigma_a^2}+\xi^2\right]\right)^{\frac{1}{2}} \tag{1.23}$$

Inserting (1.19), (1.22), and (1.23) into (1.18) we have

$$\begin{aligned}
\mathrm{E}\left[|\boldsymbol{\xi}|\,\middle|\,|\xi|<\rho_1\right] &= \frac{\int_{-\rho_1}^{\rho_1} E\left[|\boldsymbol{\xi}|\,\middle|\xi\right] f\left(\xi\right)d\xi}{\int_{-\rho_1}^{\rho_1} f\left(\xi\right)d\xi} \\
&= \frac{1}{\rho_1}\frac{\sigma_c^2}{\sigma_a^2}\left[\frac{\sigma_c^2}{\sigma_a^2}+\rho_1^2\right]^{\frac{1}{2}}\frac{4}{\pi}\int_0^{\rho_1}\frac{\sigma_c^2/\sigma_a^2\xi}{2\left[\sigma_c^2/\sigma_a^2+\xi^2\right]^{\frac{3}{2}}}d\xi \\
&= \frac{1}{\rho_1}\left(\frac{\sigma_c}{\sigma_a}\left[\frac{\sigma_c^2}{\sigma_a^2}+\rho_1^2\right]^{\frac{1}{2}}-\frac{\sigma_c^2}{\sigma_a^2}\right)
\end{aligned} \tag{1.24}$$

$$
\begin{aligned}
\mathrm{E}\left[|\boldsymbol{\tau}| \,\middle|\, |\xi| < \rho_1\right] &= \frac{\int_{-\rho_1}^{\rho_1} E\left[|\boldsymbol{\tau}| \,\middle|\, \xi\right] f\left(\xi\right) d\xi}{\int_{-\rho_1}^{\rho_1} f\left(\xi\right) d\xi} \\
&= \frac{1}{\rho_1}\frac{\sigma_c^2}{\sigma_a^2}\left[\frac{\sigma_c^2}{\sigma_a^2} + \rho_1^2\right]^{\frac{1}{2}} \frac{4}{\pi}\int_0^{\rho_1} \frac{\left(\frac{|K|}{\sigma_a^2\sigma_c^4}\left[\frac{\sigma_c^2}{\sigma_a^2} + \xi^2\right]\right)^{\frac{1}{2}} \frac{\sigma_c^2}{\sigma_a^2}}{2\left[\frac{\sigma_c^2}{\sigma_a^2} + \xi^2\right]^{\frac{3}{2}}} d\xi \\
&= \frac{2}{\pi\rho_1}\left(\frac{|K|}{\sigma_a^4\sigma_c^2}\right)^{\frac{1}{2}}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_1^2\right)^{\frac{1}{2}} \tan^{-1}\left(\frac{\sigma_a}{\sigma_c}\rho_1\right) \qquad (1.25)
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}\left[|\boldsymbol{\nu}| \,\middle|\, |\xi| < \rho_1\right] &= \frac{\int_{-\rho_1}^{\rho_1} E\left[|\boldsymbol{\nu}| \,\middle|\, \xi\right] f\left(\xi\right) d\xi}{\int_{-\rho_1}^{\rho_1} f\left(\xi\right) d\xi} \\
&= \frac{1}{\rho_1}\frac{\sigma_c^2}{\sigma_a^2}\left[\frac{\sigma_c^2}{\sigma_a^2} + \rho_1^2\right]^{\frac{1}{2}} \frac{4}{\pi}\int_0^{\rho_1} \frac{\left(\xi^2\left[\sigma_c^2/\sigma_a^2 + \xi^2\right]\right)^{\frac{1}{2}} \frac{\sigma_c^2}{\sigma_a^2}}{2\left[\sigma_c^2/\sigma_a^2 + \xi^2\right]^{\frac{3}{2}}} d\xi \\
&= \frac{1}{\pi\rho_1}\frac{\sigma_c^2}{\sigma_a^2}\left(\frac{\sigma_a^2}{\sigma_c^2} + \rho_1^2\right)^{\frac{1}{2}} \ln\left(1 + \frac{\sigma_c^2}{\sigma_a^2}\rho_1^2\right) \qquad (1.26)
\end{aligned}
$$

Finally, we would like to quantify the influence of constraint (3.13) on $\boldsymbol{\xi}$, $\boldsymbol{\tau}$, $\boldsymbol{\nu}$ again by evaluating $\mathrm{E}\left[|\boldsymbol{\xi}| \,\middle|\, |R_1| < \rho_3\right]$, $\mathrm{E}\left[|\boldsymbol{\tau}| \,\middle|\, |R_1| < \rho_3\right]$, $\mathrm{E}\left[|\boldsymbol{\nu}| \,\middle|\, |R_1| < \rho_3\right]$, where $R_1 = \sqrt{\xi^2 + \chi^2}$. We find the expected values in a slightly different manner than that shown in (1.18). In Appendix 1.2 we discussed how $\boldsymbol{\xi}$, $\boldsymbol{\tau}$, $\boldsymbol{\nu}$ can be expressed as functions of the random variables $\boldsymbol{R_1}$, $\boldsymbol{R_2}$, $\boldsymbol{\Phi_1}$ and $\boldsymbol{\Phi_2}$. (See equations (1.9), (1.11), and (1.12)). Therefore, we can use the pdfs in

(1.14) and (1.15) to calculate the expected value as follows:

$$\begin{aligned}
\mathrm{E}\big[|\boldsymbol{\xi}|\,\big|\,|R_1| < \rho_3\big] &= \frac{\int_{-\pi}^{\pi}\int_0^{\rho_3} \xi f\left(R_1\right) f\left(\Phi_1\right) dR_1 d\Phi_1}{\int_0^{\rho_3} f\left(R_1\right) dR_1} \\
&= \frac{1}{\rho_3^2}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right)\int_{-\pi}^{\pi}\int_0^{\rho_3} \frac{R_1^2\left|\sin\Phi_1\right|\sigma_c^2/\sigma_a^2}{\pi\left(\sigma_c^2/\sigma_a^2 + r^2\right)^2} dR_1 d\Phi_1 \\
&= \frac{1}{\rho_3^2}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right)\frac{2}{\pi}\left[-\cos\Phi_1\right]_0^{\frac{\pi}{2}}\int_0^{\rho_3}\frac{2R_1^3\sigma_c^2/\sigma_a^2}{\left(\sigma_c^2/\sigma_a^2 + r^2\right)^2} dR_1 \\
&= \frac{2}{\pi\rho_3}\frac{\sigma_c^2}{\sigma_a^2}\left(\frac{1}{\rho}\frac{\sigma_a}{\sigma_c}\tan^{-1}\left[\frac{\sigma_a}{\sigma_c}\rho_3\right]\left[\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right] - 1\right). \qquad (1.27)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\big[|\boldsymbol{\tau}|\big||R_1| < \rho_3\big] &= \frac{\int_{-\pi}^{\pi}\int_0^{\rho_3}\int_0^{\infty} \tau f\left(R_1, R_2\right) f\left(\Phi_2\right) dR_2 dR_1 d\Phi_2}{\int_0^{\rho_3} f\left(R_1\right) dR_1} \\
&= \frac{1}{\rho_3^2}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right)\int_{-\pi}^{\pi}\int_0^{\rho_3}\int_0^{\infty}\frac{\sigma_s^4}{\sigma_a^2\sigma_c^2}\frac{8R_1R_2^2\left|\sin\Phi_1\right|}{\frac{\sigma_s^2}{\sigma_c^2}\left[\frac{\sigma_c^2}{\sigma_a^2} + R_1^2\right] + R_2^2} dR_2 dR_1 d\Phi_2 \\
&= \frac{1}{\rho_3^2}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right)\frac{2}{\pi}\left[-\cos\Phi_1\right]_0^{\frac{\pi}{2}}\int_0^{\rho_3}\frac{\pi}{2}\frac{\sigma_s\sigma_c}{\sigma_a^2}\frac{R_1}{\left(\sigma_c^2/\sigma_a^2 + R_1^2\right)^{\frac{3}{2}}} dR_1 \\
&= \frac{1}{\rho_3^2}\frac{\sigma_c}{\sigma_a}\left(\frac{|K|}{\sigma_a^4\sigma_c^2}\right)^{\frac{1}{2}}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right)^{\frac{1}{2}}\left(\frac{\sigma_a}{\sigma_b}\left[\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right] - 1\right)^{\frac{1}{2}} \qquad (1.28)
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}\big[|\boldsymbol{\nu}|\,\big|\,|R_1| < \rho_3\big] &= \frac{\int_{-\pi}^{\pi}\int_0^{\rho_3} \nu f\left(R_1\right) f\left(\Phi_1\right) dr d\Phi_1}{\int_0^{\rho_3} f\left(R_1\right) dR_1} \\
&= \frac{1}{\rho_3^2}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right)\int_{-\pi}^{\pi}\int_0^{\rho_3}\frac{R_1^3\left|\cos\Phi_1\sin\Phi_1\right|\sigma_c^2/\sigma_a^2}{\pi\left(\sigma_c^2/\sigma_a^2 + r^2\right)^2} dR_1 d\Phi_1 \\
&= \frac{1}{\rho_3^2}\left(\frac{\sigma_c^2}{\sigma_a^2} + \rho_3^2\right)\frac{1}{\pi}\left[\sin^2\Phi_1\right]_0^{\frac{\pi}{2}}\int_0^{\rho_3}\frac{2R_1^3\sigma_c^2/\sigma_a^2}{\left(\sigma_c^2/\sigma_a^2 + r^2\right)^2} dR_1 \\
&= \frac{\sigma_c^2/\sigma_a^2}{\pi}\left[\frac{\sigma_c^2/\sigma_a^2 + \rho_3^2}{\rho_3^2}\ln\left(\frac{\sigma_c^2/\sigma_a^2 + \rho_3^2}{\sigma_c^2/\sigma_a^2}\right) - 1\right]. \qquad (1.29)
\end{aligned}$$

Bibliography

[1] M. Agrawal and L. S. Davis. Window-based, discontinuity preserving stereo. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:66–73, 2004.

[2] U. Ahlvers and U. Zoelzer. Improvement of phase-based algorithms for disparity estimation by means of magnitude information. *Proceedings of the IEEE International Conference on Image Processing*, 2:3025–3028, 2004.

[3] U. Ahlvers, U. Zoelzer, and S. Rechmeier. Fft-based disparity estimation for stereo image coding. *Proceedings of the IEEE International Conference on Image Processing*, 1:761–764, 2003.

[4] J. Aloimonos, I. Weiss, and A. Bandyopadhyay. Active vision. *Proceedings of the First International Conference on Computer Vision*, pages 35–54, 1987.

[5] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3):283–310, January 1989.

[6] A. Anzai, I. Ohzawa, and R. Freeman. Neural mechanisms for processing binocular information i. simple cells. *Journal of Neurophysiology*,

82(2):891–908, Aug 1999.

[7] C. Bandera and P. D. Scott. Foveal machine vision systems. *Proceedings of the International Conference on Systems, Man and Cybernetics*, 2:596–599, 1989.

[8] M. S. Banks, Sergei G., and H. F. Rose. Local cross-correlation model of stereo correspondence. *Proceedings of SPIE - The International Society for Optical Engineering*, 5666:53–61, 2005.

[9] A. Basu. Optimal discretization for stereo reconstruction. *Pattern Recognition Letters*, 13(11):813–820, November 1992.

[10] A. Basu and H. Sahabi. Optimal non-uniform discretization for stereo reconstruction. *Proceedings of the IEEE International Conference on Pattern Recognition*, 1:755–759, 1996.

[11] A. Basu and H. Sahabi. Analysis of depth estimation error for cylindrical stereo imaging. *Pattern Recognition*, 35(11):2549–2558, 2002.

[12] A. Bernardino and J. Santos-Victor. Vergence control for robotic heads using log-polar images. *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, 3:1264–1271, 1996.

[13] A. Bernardino and J. Santos-Victor. A binocular stereo algorithm for log-polar foveated systems. In *Proceedings of the Second International Workshop on Biologically Motivated Computer Vision*, pages 127–136, London, UK, 2002. Springer-Verlag.

[14] C. Blakemore. The range and scope of binocular depth discrimination in man. *Journal of Physiology*, 211(3):599–622, Dec 1970.

[15] A. C. Bovik. *The Handbook of Image & Video Processing*. Academic Press, San Diego, CA, 2000.

[16] A. C. Bovik, M. Clark, and W. S. Geisler. Computational texture analysis using localized spatial filtering. *Proceedings of the Workshop on Computer Vision*, pages 201–206, 1987.

[17] A. C. Bovik, M. Clark, and W. S. Geisler. Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):55–73, 1990.

[18] L.-D. Cai and J. Mayhew. Estimating mean disparity of stereo images using shift-trials of phase differences. *Proceedings of the British Machine Vision Association*, pages 237–246, 1992.

[19] L.-D. Cai and J. Mayhew. Note on some phase differencing algorithms for disparity estimation. *International Journal of Computer Vision*, 22(2):111–124, 1997.

[20] C. Capurro, F. Panerai, and G. Sandini. Dynamic vergence using log-polar images. *International Journal of Computer Vision*, 24(1):79–94, 1997.

[21] G. Carneiro and A. D. Jepson. Multi-scale phase-based local features. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:736–743, 2003.

[22] G. Carneiro and A. D. Jepson. The distinctiveness, detectability, and robustness of local image features. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2:296–301, 2005.

[23] Q.-S. Chen and F. Deconinck. Foveal-view stereopsis using symmetric phase-only matched filtering. *Proceedings of SPIE - The International Society for Optical Engineering*, 2233:46–53, 1994.

[24] T.-Y. Chen and A. C. Bovik. Stereo vision using gabor wavelets. *IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 13–17, April 1994.

[25] T.-Y. Chen and A. C. Bovik. Stereo disparity from multiscale processing of local image phase. *Proceedings of the IEEE International Conference on Computer Vision*, pages 188–193, 1995.

[26] T.-Y. Chen, A. C. Bovik, and L. K. Cormack. Stereoscopic ranging by matching image modulations. *IEEE Transactions on Image Processing*, 8(6):785–797, 1999.

[27] T.-Y. Chen, A. C. Bovik, and B. J. Super. Multiscale stereopsis via gabor filter phase response. *Proceedings of theIEEE International Conference on Systems, Man and Cybernetics*, 1:55–60, 1994.

[28] J. Clark, M. Palmer, and P. Lawrence. A transformation method for the reconstruction of functions from nonuniformly spaced samples. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 33(5):1151–1165, Oct 1985.

[29] M. Clark and A. C. Bovik. Texture discrimination using a model of the visual cortex. *Procedings of the IEEE International Conference on Systems, Man, and Cybernetics*, pages 1425–1430, 1986.

[30] M. Clark and A. C. Bovik. Experiments in segmenting texton patterns using localized spatial filters. *Pattern Recognition*, 22(6):707–717, 1989.

[31] M. Clark, A. C. Bovik, and W. Geisler. Texture segmentation using a class of narrowband filters. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 12, pages 571–574, Apr 1987.

[32] L. K. Cormack, Y. Liu, and A. C. Bovik. Disparity statistics in the natural environment. Vision Sciences Society 5th Annual Meeting, May 2005.

[33] L. K. Cormack and R. B. Riddle. Binocular correlation detection with oriented dynamic random line stereograms. *Vision Research*, 36(15):2303–2310, Aug 1996.

[34] L. K. Cormack, S. B. Stevenson, and C. M. Schor. An upper limit to the binocular combination of stimuli. *Vision Research*, 34(19):2599–2608,

Oct 1994.

[35] A. Cozzi, B. Crespi, F. Valentinotti, and F. Woergoetter. Performance of phase-based algorithms for disparity estimation. *Machine Vision and Applications*, 9(5-6):334–340, 1997.

[36] B. Crespi and G. Tecchiolli. Adaptive gabor filters for phase-based disparity estimation. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(5):591–614, 1999.

[37] J. Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20(10):847–856, 1980.

[38] G. C. DeAngelis, I. Ohzawa, and R. D. Freeman. Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature*, 352(6331):156–159, July 1991.

[39] U. R. Dhond and J.K. Aggarwal. Structure from stereo - a review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, 1989.

[40] L. Di Stefano, M. Marchionni, S. Mattoccia, and G. Neri. Quantitative evaluation of area-based stereo matching. *Proceedings of the IEEE International Conference on Control, Automation, Robotics and Vision*, 2:1110–1115, 2002.

[41] A. El Zaart, D.l Ziou, and F. Dubeau. Phase-based disparity estimation: A spatial approach. *Porceedings of the the IEEE International Conference on Image Processing*, 3:244–247, 1997.

[42] A. Elnagar. Optimal error discretization under depth and range constraints. *Pattern Recognition Letters*, 19(9):879–888, 1998.

[43] J. T. Enright. Stereopsis, cyclotorsional "noise" and the apparent vertical. *Vision Research*, 30(10):1487–1497, 1990.

[44] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M. Zhang. Foveated visual tracking chip. *Digest of Technical Papers - IEEE International Solid-State Circuits Conference*, 40:38–39, 1997.

[45] R. Etienne-Cummings, J. Van der Spiegel, P. Mueller, and M. Zhang. Foveated silicon retina for two-dimensional tracking. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, 47(6):504–517, 2000.

[46] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera self calibration: Theory and experiments. *Proceedings of the Second European Conference on Computer Vision*, 588:321–334, 1992.

[47] D. Fleet. Disparity from local weighted phase-correlation. *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, 1:48–54, 1994.

[48] D. Fleet and A. Jepson. Stability of phase information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(12):1253–1268, 1993.

[49] D. Fleet, A. Jepson, and M. Jenkin. Phase-based disparity measurement. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 53(2):198–210, 1991.

[50] J. M. Foley, T. H. Applebaum, and W. A. Richards. Stereopsis with large disparities: discrimination and depth magnitude. *Vision Research*, 15(3):417–421, Mar 1975.

[51] T. Frohlinghaus and J. Buhmann. Regularizing phase-based stereo. In *Proceedings of the 13th International Conference on Pattern Recognition*, volume 1, pages 451–455vol.1, 25-29 Aug. 1996.

[52] D. Gabor. Theory of communication. *Journal Institute of Electrical Engineers*, 93:429–441, 1946.

[53] W. S. Geisler. Vision systems: Course notes. January 2001.

[54] W. E. Grimson. A computer implementation of a theory of human stereo vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 292(1058):217–253, May 1981.

[55] J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Yale University Press, 1923.

[56] B. Horn. Non-correlation methods for stereo matching. *Photogrammetric Engineering and Remote Sensing*, 49:535–536, 1983.

[57] B. Horn. *Robot Vision*. The MIT Press, 1986.

[58] I.P. Howard and B.J. Rogers. *Binocular Vision and Stereopsis*. Oxford University Press, 1995.

[59] T. Jaworska. Multiresolution wavelet analysis used to stereo image matching. *Proceedings of the International Conference on Systems Science*, 2:85–92, 2001.

[60] T. Jaworska. Amplitude elimination for stereo image matching based on the wavelet approach. *Machine Graphics and Vision*, 14(1):103–120, 2005.

[61] M. Jenkin and A. Jepson. Recovering local surface structure through local phase difference measurements. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 59(1):72–93, 1994.

[62] M. Jenkin, A. Jepson, and J. Tsotsos. Techniques for disparity measurement. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 53(1):14–30, 1991.

[63] A. Jepson and D. Fleet. Phase singularities in scale-space. *Image and Vision Computing*, 9(5):338–343, 1991.

[64] A. D. Jepson and M. Jenkin. Fast computation of disparity from phase differences. *Proceedings of the IEEE Computer Vision and Pattern Recognition*, pages 398–403, 1989.

[65] J.P. Jones and L.A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58:1233–1258, 1987.

[66] J.P. Jones and L.A. Palmer. The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58:1187–1211, 1987.

[67] III Jordan, J.R. and A.C. Bovik. Computational stereo vision using color. *IEEE Control Systems Magazine*, 8(3):31–36, June 1988.

[68] J. R. Jordan, W. S. Geisler, and A. C. Bovik. Color as a source of information in the stereo correspondence process. *Vision Research*, 30(12):1955–1970, 1990.

[69] J.-G. Kim and H. Jeong. Parallel relaxation algorithm for disparity computation. *Electronics Letters*, 33(16):1367–1368, 1997.

[70] W. N. Klarquist and A. C. Bovik. Fovea: A foveated vergent active stereo vision system for dynamic three-dimensional scene recovery. *IEEE Transactions on Robotics and Automation*, 14(5):755–770, 1998.

[71] F. Li and Y. Zhou. Disparity estimation based on frequency domain. *Shanghai Jiaotong Daxue Xuebao/Journal of Shanghai Jiaotong University*, 33(5):516–519, 1999.

[72] Z.-N. Li and F. Tong. Reciprocal-wedge transform in active stereo. *International Journal of Pattern Recognition and Artificial Intelligence*, 13(1):25–48, 1999.

[73] J. Magarey and A. Dick. Multiresolution stereo image matching using complex wavelets. *Proceedings of the 14th International Conference on Pattern Recognition*, 1:4–7, August 1998.

[74] R. Manzotti, A. Gasteratos, G. Metta, and G. Sandini. Disparity estimation on log-polar images and vergence control. *Computer Vision and Image Understanding*, 83(2):97–117, 2001.

[75] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194(4262):283–287, Oct 1976.

[76] D. Marr and T. Poggio. A computational theory of human stereo vision. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 204(1156):301–328, May 1979.

[77] J. Mayhew and J. Frisby. Convergent disparity discriminations in narrow band filtered random dot stereograms. *Vision Research*, 19(1):63–71, 1979.

[78] K. Nishino and S. K. Nayar. The world in an eye. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:444–451, 2004.

[79] K. Nishino and S. K. Nayar. Corneal imaging system: Environment from eyes. *International Journal of Computer Vision*, 70(1):23–40, 2006.

[80] I. Ohzawa. Mechanisms of stereoscopic vision: the disparity energy model. *Current Opinion in Neurobiology*, 8(4):509–515, Aug 1998.

[81] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman. Encoding of binocular disparity by complex cells in the cat's visual cortex. *Journal of Neurophysiology*, 77(6):2879–2909, Jun 1997.

[82] I. Ohzawa, G. C. DeAngelis, and R. D. Freeman. The neural coding of stereoscopic depth. *Neuroreport*, 8(3):3–12, Feb 1997.

[83] I. Ohzawa and R. D. Freeman. The binocular organization of complex cells in the cat's visual cortex. *Journal of Neurophysiology*, 56(1):243–259, Jul 1986.

[84] I. Ohzawa and R. D. Freeman. The binocular organization of simple cells in the cat's visual cortex. *Journal of Neurophysiology*, 56(1):221–242, Jul 1986.

[85] P. Olver and C.i Shakiban. *Fundamentals of Applied Mathematics.* Prentice-Hall, in preparation.

[86] A.V. Oppenheim and J.S. Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, May 1981.

[87] M. Ouali, D. Ziou, and C. Laurgeau. Cooperative multiscale phase-based disparity algorithm. *Proceedings of the IEEE International Conference on Image Processing*, 3:145–149, 1999.

[88] M. Ouali, D. Ziou, and C. Laurgeau. Dense disparity estimation using gabor filters and image derivatives. In *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on*, pages 483–489, 4-8 Oct. 1999.

[89] M.H. Ouali, H. Lange, and C. Laurgeau. Energy minimization approach to dense stereovision. *Proceedings of the IEEE International Conference on Image Processing*, 2:841–845, 1996.

[90] M.H. Ouali, D. Ziou, and C. Laurgeau. Inaccurate phase-based disparities removal. *Proceedings of International Conference on Electronics, Circuits, and Systems*, 2:5–8, September 1999.

[91] M.H. Ouali, D. Ziou, and C. Laurgeau. Using gabor decomposition to improve dense disparity estimation. *Proceedings of SPIE - The International Society for Optical Engineering*, 3808:634–644, 1999.

[92] A. Palacios and J. Plazas. A foveal architecture for stereo matching. *Proceedings of the IEEE International Conference on Image Processing*, 2:521–524, 2002.

[93] H. Pan and J. Magarey. Multiresolution phase-based bidirectional stereo matching with provision for discontinuity and occlusion. *Digital Signal Processing: A Review Journal*, 8(4):255–266, 1998.

[94] D.V. Papadimitriou and T.J. Dennis. Stereo disparity analysis using phase correlation. *Electronics Letters*, 30(18):1475–1477, 1994.

[95] D.V. Papadimitriou and T.J. Dennis. Stereo disparity algorithm for 3d model construction. *Fifth International Conference on Image Processing and its Applications*, (410):178 –182, 1995.

[96] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, Inc., New York, NY, 1991.

[97] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(6035):314–319, 1985.

[98] H. Sahabi and A. Basu. Analysis of error in depth perception with vergence and spatially varying sensing. *Computer Vision and Image Understanding*, 63(3):447–461, 1996.

[99] T.D. Sanger. Stereo disparity computation using gabor filters. *Biological Cybernetics*, 59:405–418, 1988.

[100] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 195–202, 2003.

[101] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 131–140, 2001.

[102] K. Schindler. Geometry and construction of straight lines in log-polar images. *Computer Vision and Image Understanding*, 103(3):196–207, 2006. ;.

[103] K. Schindler and H. Bischof. The epipolar geometry of the log-polar image plane. *Proceedings of the International Conference on Pattern Recognition*, 4:40–43, 2004.

[104] B.Y. Schlesinger and Y. Yeshurun. Spatial size limits in stereoscopic vision. *Spatial Vision*, 11(3):279–293, 1998.

[105] E. L. Schwartz. Spatial mapping in the primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25(4):181–194, Feb 1977.

[106] E. L. Schwartz. Computational anatomy and functional architecture of striate cortex: a spatial mapping approach to perceptual coding. *Vision Research*, 20(8):645–669, 1980.

[107] E. L. Schwartz. A quantitative model of the functional architecture of human striate cortex with application to visual illusion and cortical texture analysis. *Biological Cybernetics*, 37(2):63–76, 1980.

[108] E. L. Schwartz, D. N. Greve, and G. Bonmassar. Space-variant active vision: definition, overview and examples. *Neural Networks*, 8(7-8):1297–1308, 1995.

[109] E. L. Schwartz, B. Merker, E. Wolfson, and A. Shaw. Applications of computer graphics and image processing to 2d and 3d modeling of the functional architecture of visual cortex. *IEEE Computer Graphics and Applications*, 8(4):13–23, 1988.

[110] S. Shah and J.K. Aggarwal. Depth estimation using stereo fish-eye lenses. *Proceedings of the IEEE International Conference on Image Processing*, 2:740–744, 1994.

[111] S.-W. Shih, Y.-P. Hung, and W.-S. Lin. Calibration of an active binocular head. *IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans*, 28(4):426–442, 1998.

[112] F. Solari, S.P. Sabatini, and G.M. Bisio. Fast technique for phase-based disparity estimation with no explicit calculation of phase. *Electronics Letters*, 37(23):1382–1383, 2001.

[113] S. B. Stevenson and C. M. Schor. Human stereo matching is not restricted to epipolar lines. *Vision Research*, 37(19):2717–2723, Oct 1997.

[114] M. Tistarelli and G. Sandini. Direct estimation of time-to-impact from optical flow. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 226–233, 7-9 Oct. 1991.

[115] I. Ulusoy, U. Halici, and E. R. Hancock. Probabilistic phase based sparse stereo. *Proceedings of the International Conference on Pattern Recognition*, 4:84–87, 2004.

[116] F. Valentinotti and S. Taraglio. Phase difference stereo disparity computation on a simd parallel machine. *Lecture Notes in Computer Science*, 1225:127–136, 1997.

[117] R. van Ee and C. M. Schor. Unconstrained stereoscopic matching of lines. *Vision Research*, 40(2):151–162, 2000.

[118] S. Venkatesh, J. Cooper, and B. White. Local energy and pre-envelope. *Pattern Recognition*, 28(8):1127–1134, 1995.

[119] B. Voelpel and W.M. Theimer. Localization uncertainty in area-based stereo algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, 25(12):1628–1634, 1995.

[120] S.-C. Wang and C.-L. Huang. Structure recovery and motion estimation from stereo motion. *Institute of Electronics, Information, and Communication Engineers Transactions on Information and Systems*, E77-D(11):1247–1258, 1994.

[121] Z. Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, March 2002.

[122] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *Conference Record of the Thirty-*

Seventh Asilomar Conference on Signals, Systems and Computers, volume 2, pages 1398–1402Vol.2, 9-12 Nov. 2003.

[123] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E.-H. Yang, and A.C. Bovik. Quality-aware images. *IEEE Transactions on Image Processing*, 15(6):1680–1689, June 2006.

[124] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, April 2004.

[125] J. Wei and Z.-N. Li. Efficient disparity-based gaze control with foveate wavelet transform. *IEEE International Conference on Intelligent Robots and Systems*, 2:866–871, 1998.

[126] C. F. Weiman. Efficient discrete gabor functions for robot vision. *Proceedings of SPIE - The International Society for Optical Engineering*, 2242:148–160, 1994.

[127] C. F. Weiman. Binocular stereo via log-polar retinas. *Proceedings of SPIE - The International Society for Optical Engineering*, 2488:309–320, 1995.

[128] C. F. Weiman and G. Chaikin. Logarithmic spiral grids for image processing and display. *Computer Graphics and Image Processing*, 11(3):197–226, 1979.

[129] J. Weng. Windowed fourier phase. completeness and signal reconstruction. *IEEE Transactions on Signal Processing*, 41(2):657–666, 1993.

[130] F. Woergoetter and A. Cozzi. Computing stereoscopic disparity with binocular cortical simple and complex cells. *IEEE Conference Publication*, 1(470):269–273, 1999.

[131] Y. Xu, L. Du, Zi. Hou, and G. Jin. Grouped scale-adaptive phase-based stereo matching method. *International Conference on Signal Processing Proceedings*, 2:893–896, 1998.

[132] Y. Xu, L. Du, Zi. Hou, and G. Jin. Scale-adaptive phase-based stereo matching method. *Tien Tzu Hsueh Pao/Acta Electronica Sinica*, 27(7):38–41, 1999.

[133] Y. Xu, J. Zhou, and G. Zhai. 2d phase-based matching in uncalibrated images. In *IEEE Workshop on Signal Processing Systems Design and Implementation*, pages 325–330, 2005.

[134] Y.Y. Zeevi and E. Shlomot. Nonuniform sampling and antialiasing in image representation. *IEEE Transactions on Signal Processing*, 41(3):1223–1236, March 1993.

[135] Z. Zhang, Q.-T. Luong, and O. Faugeras. Motion of an uncalibrated stereo rig: Self-calibration and metric reconstruction. *Proceedings of the International Conference on Pattern Recognition*, 1:695–697, 1994.

[136] Huaibin Zhao and J. K. Aggarwal. 3d reconstruction of an urban scene from synthetic fish-eye images. In *Proceedings of the 4th IEEE Southwest Symposium on Image Analysis and Interpretation*, page 219, Washington, DC, USA, 2000. IEEE Computer Society.

Vita

James Peter Monaco is the son of James and Linda Monaco. He received a B.S. in Computer Engineering and an M.S. in Electrical Engineering from Texas A&M University in 1996 and 1998, respectively. He has worked for Texas Instruments Inc. and Raytheon Corporation designing infrared tracking systems for military aircraft. He is currently working for VuCOMP, a medical imaging company.

Permanent address: 2300 Kathryn Lane #2223
Plano, Texas 75025

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.